Production des biens communs numériques et usages cartographiques

Flora Hayat

▶ **To cite this version:**

Flora Hayat. Production des biens communs numériques et usages cartographiques. Géographie. Université Paris Cité, 2019. Français. NNT : 2019UNIP7135 . tel-03117730

HAL Id: tel-03117730
https://theses.hal.science/tel-03117730

Submitted on 21 Jan 2021

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.

Université de Paris

École doctorale 382, Économies, Espaces, Sociétés, Civilisations : Pensée critique, politique et pratiques sociales

Laboratoire Géographie-Cités

Production des biens communs numériques et usages cartographiques

Par Flora Hayat

Thèse de doctorat de Géomatique

Dirigée par Christine Zanin

Présentée et soutenue publiquement le 30 septembre 2019

Devant un jury composé de :

Thierry Joliveau, Professeur à l'Université Jean Monet de Saint-Etienne, UMR 5600 EVS. Rapporteur
Renaud Le Goix, Professeur à l'Université Paris Diderot - Paris 7, UMR 8504 Géographie-Cités. Président
Matthieu Noucher, Chargé de recherche, équipe PASSAGE UMR 5319 Passages. Rapporteur
Ana-Maria Olteanu-Raimond, Chargée de recherche, IGN équipe COGIT. Examinatrice
Marta Severo, Professeure à l'Université Paris Nanterre– Paris X, EA 7339 laboratoire Dicen-IDF. Examinatrice
Christine Zanin, Maître de conférence, HDR à l'Université Paris Diderot – Paris 7, UMR 8504 Géographie-Cités. Directrice

Except where otherwise noted, this is work licensed under
https://creativecommons.org/licenses/by-nc-nd/3.0/fr/

Remerciements

J’adresse mes plus vifs remerciements à Christine Zanin pour avoir dirigé ce travail avec attention, disponibilité et bienveillance. Je lui suis sincèrement reconnaissante pour la confiance qu’elle m’a accordée durant ces quatre années, m’encourageant toujours à explorer de nouvelles pistes de recherche. J’espère pouvoir poursuivre, à l’avenir, les débats.

Je remercie les institutions qui ont permis la réalisation de cette thèse qui a été financée en partie par Michelin et en partie par le ministère de l’Industrie grâce à une bourse CIFRE et a été réalisé au sein de l’Université Paris 7 et de l’UMR Géographie-cités.

Je remercie également Thierry Joliveau, Renaud Le Goix, Matthieu Noucher, Ana-Maria Olteanu-Raimond, Marta Severo d’avoir accepté d’être les membres de mon jury.

J’adresse ma gratitude à l’ensemble de l’équipe du service de production cartographique pour son accueil et sa convivialité, tout particulièrement à Ludovic, Gaëlle, Laurence, Phillipe, Philippe, Sandrine et à Vincent qui m’ont fait bénéficier de leur expérience.

Au sein de l’équipe cartographie de Michelin comme dans l’équipe enseignante et celles des chercheurs à l’Université Paris Diderot, j’ai eu la chance de côtoyer des collègues et amis durant trois ans. Les principes de bienveillance, de cohésion d’équipe et de confiance dans le travail sont partagés en ces deux lieux d’accueils. Ce sont ces points communs fondamentaux qui m’ont permis de naviguer pendant trois ans d’un bureau à l’autre tous les deux jours et demi. Dans ces contextes de travail et d’amitié je remercie particulièrement :

- Les bibs-contributeurs à OSM, qui ont su se saisir d’OSM et y exprimer tous leurs talents de cartographes. Merci à Gaëlle, Philipe Co., Philipe Che, Philippe Ca., Sandrine, Laurence, Virginie, Yanzhuo pour votre patience et votre humour face aux difficultés que représente une chaîne de production en construction.
- mes voisins directs de bureau Gaëlle et Romain pour leur goût partagé du travail à voix haute en écoutant, lorsque cela est nécessaire, les *tubes* des années 1980.
- les membres de La Coordonnerie pour nos échanges passionnés.
- les habitants et voisins du bureau 768, chacun présentant des qualités originales dont la capacité de fédérer, l’humour, le panache, des talents de conteur, le goût du risque, l’entrain, l’énergie, l’élégance, le rire, la sérénité, la créativité ou la force.
- Mes amis et les experts les plus sérieux que sont Chloé et Matthieu.
- Les relecteurs, si attentionnés et généreux que sont Chloé, Chloé, Etienne, Jeannine, Marie-Coquet, Matthieu, Pierre et Romain.

Cette thèse repose également sur les contributeurs à OpenStreetMap, à Wikipédia et à Linux et sur les concepteurs et bâtisseurs de l’infrastructure d’Internet.

Si cette thèse est le résultat d’heures de travail réalisées dans les deux bureaux cités, je dois à mon entourage la stabilité et la diversité des événements et des expériences qui m’ont permis d’achever cette thèse. Pour cela, merci à Claire, Clément, Chloé, Chloé, Daphné, Dorian, Émile, Florence, Florian, Jeannine, Joséphine, Julie, Luc, Lucie, Marie-Coquet, Matthieu, Noémie, Pascal, Pierre, Romain, Vincent, Violaine et tant d’autres.

Sommaire

Avant-propos 1

INTRODUCTION GÉNÉRALE 3

PARTIE I : Informations géographiques à l'ère du numérique 13

Chapitre 1 : Changements de paradigme en cartographie 17

Conclusion du chapitre 1 99

Chapitre 2 : Fiabilité de la base de données OpenStreetMap 103

Conclusion du chapitre 2 171

PARTIE I. Conclusion : OSM une source unificatrice ? 175

PARTIE II : Cartographie industrielle et communs numériques 179

Chapitre 3 : Entre publications cartographiques commerciales et productions libres d'informations géographiques 183

Conclusion du chapitre 3 285

Chapitre 4 : Quel avenir pour l'Information géographique libre ? 289

Conclusion du chapitre 4 332

PARTIE II. Conclusion 333

CONCLUSION GÉNÉRALE 337

Annexes 347

Avant-propos

Ce travail de thèse s'insère directement dans le cadre d'un contrat CIFRE signé avec l'entreprise Michelin, pour une période de 3 ans, entre mars 2016 et mars 2019. Ce contrat avait pour objectif la réalisation d'une recherche appliquée pour la commercialisation de cartographies touristiques et routières. Le projet visait à évaluer les apports techniques, méthodologiques et idéologiques du mouvement Libre à la construction d'une information géographique de qualité, nécessaire à la réalisation de cartes par une entreprise traditionnelle.

L'entreprise de cartes touristiques Michelin, leader du marché dans ce domaine, qui est installée dans un bâtiment de Boulogne-Billancourt, m'a donc accueillie dans le cadre d'une procédure CIFRE. Peu après mon arrivée, en mars 2015 à l'étage des cartographes, je me suis intégrée à une équipe soudée. Plusieurs de mes collègues conservaient le savoir et le savoir-faire cartographique développés et accumulés par l'entreprise depuis le développement touristique du réseau routier. Tel collègue connaissait par cœur le tracé de toutes les côtes françaises, telle autre était une dessinatrice minutieuse qui nuançait ses rendus de belles couleurs, tel autre conservait la mémoire de toutes les cartes déjà publiées par Michelin. À travers eux, j'ai pu mesurer la qualité du patrimoine cartographique d'une entreprise de référence.

J'ai beaucoup appris auprès de ces experts qui, souvent, ont fait toute leur carrière chez Michelin. Un travail en équipe s'est rapidement instauré autour du projet d'exploitation de la base de données OSM, en y intégrant l'apprentissage de la contribution au projet et la manipulation des données par les moyens techniques adaptés. Ces nouvelles procédures d'établissement des cartes sont utiles aux cartographes. Elle leur permet à la fois un retour à la production des informations géographiques par la contribution et un accès à des informations originales.

Les trois années de mon contrat, commencé le 15 mars 2016, correspondent à une mutation de la cartographie Michelin. Durant cette période, l'entreprise a élaboré certaines de ses nouvelles cartes à partir des données OSM. Mon passage à Boulogne-Billancourt a ainsi été concomitant avec une période charnière de l'histoire de

l'entreprise, au cours de laquelle les cartographes ont eu la possibilité de tester l'association des techniques traditionnelles avec les techniques du Libre. C'est la raison pour laquelle l'entreprise traditionnelle Michelin, devenue un acteur des usages cartographiques du Libre à travers son service cartographique, servira souvent d'exemple dans notre analyse.

Je fais le choix de construire cette recherche à l'intersection de plusieurs disciplines telles que la géographie, la géomatique, la sociologie, l'économie, la modélisation la classification des données et le droit. La démarche que nous proposons nous permet d'explorer la nouveauté des sources d'informations libres en considérant qu'elles sont produites dans un écosystème numérique dynamique. Ces regards croisés ont relevé pour nous de l'observation distanciée et de l'analyse théorique. Ils nous ont également permis d'interroger nos propres pratiques géomatiques. Afin de faciliter la compréhension de cet écosystème avec la prise de distance adaptée, je présente mon travail de recherche en adoptant la position d'une internaute, qui a partagé durant trois ans son temps de travail entre une institution garante de méthodes de production expertes (l'université Paris Diderot et le laboratoire Géographie-cités) et l'entreprise Michelin, productrice de référence de cartes touristiques. Cette position m'a permis de conserver un point de vue réflexif, tout en évitant de valoriser outre mesure les résultats atteints par le mouvement du Libre.

INTRODUCTION GÉNÉRALE

Le sujet central de cette thèse porte sur les usages cartographiques, dans des contextes industriels de production de documents papiers ou numériques, de la base de données libre OpenStreetMap (OSM) construite *a priori*, par des bénévoles, non-experts, de la géographie.

Notre recherche procède ainsi d'un constat initial : l'usage d'une base de données produite par des contributeurs anonymes, *a priori* non experts de l'information géographique et qui ne sont tenus à aucun engagement de qualité, suscite inévitablement inquiétudes et interrogations, méfiance et défiance. Cependant, dans le même temps, cette situation nouvelle soulève un grand intérêt et semble correspondre à des attentes pressantes. La source OSM peut en effet présenter plusieurs avantages. D'une part, elle peut combler des manques thématiques et participer à de nouveaux usages cartographiques et géographiques nécessitant, par exemple, des données dont la conception serait partagée à une échelle fine selon des emprises transnationales. D'autre part, les connaissances relatives au contexte de production de la base de données OSM, à son écosystème1, aux modalités de la diffusion de ses données et de ses techniques permettent d'entrevoir les perspectives futures de l'information géographique libre et d'apporter un éclairage sur d'autres informations géographiques diffusées sous l'appellation *open data*[2] aux formats, quant à eux, très divers.

La contribution par des internautes à la base de données OpenStreetMap, dont il est question ici, a pour objectif fondamental de construire un bien commun de la connaissance de l'information géographique dont la diffusion et la protection reposent sur les clauses de la licence de distribution des données, la licence ODbL. Chacun est libre de diffuser les données, d'en produire de nouvelles ou de les modifier en y

1 Nous utilisons l'analogie de l'écosystème mobilisée dans le domaine de l'environnement afin de décrire un ensemble d'éléments en interaction dans un milieu. Cette analogie est employée pour désigner un écosystème dont le milieu est Internet et qui favorise les logiques interrelationnelles. Pour Lise Viera (2015), « le terme écosystème numérique se définit comme un ensemble dynamique composé d'éléments matériels : sites web, réseaux sociaux, plateformes, logiciels et de différents types d'acteurs : créateurs, producteurs, diffuseurs, usagers. Chacun de ces éléments est en interdépendance avec tous les autres et contribue à constituer la complexité et la richesse de cet ensemble qui génère des interactions aux équilibres subtils, parfois éphémères et en constante évolution, caractéristiques de l'écologie numérique ».

2 L'absence de traduction littérale de certains termes en anglais signifie qu'ils ne doivent pas être entendu dans leur sens commun mais comme des concepts, des catégories ou des objets spécifiques à l'écosystème de la donnée numérique contributive.

apportant des changements ou en les transformant. Cette contribution consiste en l'inscription de l'internaute au projet OpenStreetMap, via le site openstreetmap.org. Une fois inscrit, celui-ci peut procéder à l'édition de la base de données, élément central du projet OSM. En 2004, le projet OSM est conçu et proposé au public par un groupe de quelques Britanniques, passionnés de l'information géographique, et investis dans la défense de la diffusion libre des connaissances. Ce projet s'est rapidement constitué en une association britannique, la Fondation OSM. Le contributeur, inscrit sur le site openstreetmap.org, n'est cependant pas tenu de devenir membre de la Fondation, garante de l'intégrité légale de la base de données. En outre, l'internaute peut également devenir contributeur en participant à la construction d'une infrastructure technique, spécifiquement adaptée aux choix techniques et méthodologiques des contributeurs à la base de données. Ces développements ne peuvent se faire sans interactions sociales fortes. Nous observons ces relations se constituer au travers de forums et de sites de discussions. Ce sont d'ailleurs ces interactions sociales qui ont fait d'OSM un projet mondial, mobilisé par les entreprises du domaine de l'information géographique et modèle du système contributif en ligne.

Une source d'information aussi novatrice dans son système de production, qui a pour principal socle de communication le réseau Internet et pour principal ensemble technique le Web a besoin d'être étudiée dans sa complexité. Les innovations socio-techniques fondamentales, à associer au Web et à l'émergence du projet OSM, se sont manifestées plus largement dans le domaine de la géographie par ce qui est habituellement nommé le Géoweb (Joliveau, 2011). Le cartographe contemporain inscrit en effet son travail dans le nouveau contexte de production que constitue le Géoweb qui est un ensemble de moyens permettant à la fois la représentation rapide d'informations géographiques et la création de données géographiques. Cette situation nouvelle de la cartographie a des incidences directes sur sa relation complexe à l'émergence de la « nouvelle géographie ». C'est en 2006, qu'Andrew Turner publie *Introduction to Neogeography*. Cette « nouvelle géographie » serait composée de nouvelles formes d'expertises, portées par des individus étrangers à la connaissance cartographique ou géomatique, qui seraient néanmoins susceptibles de renouveler les savoirs et les pratiques géographiques et cartographiques (Turner, 2006 ; Joliveau et Rana, 2009). Nous nous sommes cependant demandé s'il ne convenait pas de nuancer cette approche et de considérer que les compétences entre experts et profanes sont

complémentaires. Il s'agira alors d'identifier les motivations des experts à participer aux projets libres et à reposer la question de l'inexpérience *a priori* du contributeur bénévole. Une base de données au modèle évolutif, créée et enrichie par une communauté non hiérarchisée, n'est pas une entité simple à analyser. Ainsi nous procéderons à l'analyse distincte des éléments qui la constituent et qui soutiennent sa pérennité. Cette analyse sera doublée de l'analyse de l'écosystème qui s'est constitué autour du projet OSM. Nous ferons intervenir trois types d'acteur, pas toujours simples à discriminer, mais que nous distinguons par souci de méthode : l'internaute, l'institution et l'entreprise privée à visée commerciale.

L'acteur principal de notre investigation n'est pas l'entreprise elle-même mais l'internaute. L'internaute est un individu qui peut être anonyme et qui est connecté au réseau Internet. Il peut choisir d'agir seul et, dans notre contexte, décider de participer, au sein d'une communauté aux relations distendues, à la base de données qu'il prévoit d'utiliser ou non. L'internaute que nous ciblons peut aisément devenir à la fois producteur et utilisateur de données. Le groupe dont il fait partie, plus ou moins organisé, est un agglomérat de contributeurs et d'utilisateurs qui ont en commun le dessein de développer et de maintenir la base de données OpenStreetMap. Sachant que tout le groupe défend le projet OpenStreetMap. Il peut, s'il le souhaite, rejoindre un groupe d'internautes réunis par des intérêts communs et poursuivre les objectifs de ce groupe, tout en continuant à servir ses propres intérêts. En retour, le groupe l'aide à atteindre ses buts. L'internaute peut donc intervenir seul ou collectivement au sein même d'un groupe. Le système de production de la base de données OSM est donc maintenu par des internautes-contributeurs et il peut être utilisé par n'importe quel internaute, même s'il est non producteur et inconnu du groupe.

Le deuxième type d'acteur est à rechercher du côté des institutions. Les institutions qui exercent une autorité dans les domaines de l'information géographique, des données et des protocoles d'échanges de données sont les garantes d'une production contributive synchronisée. En effet, le système de production de la base de données OSM s'inscrit dans un contexte international. Les protocoles du réseau Internet sont, en principe, ouverts à tous et le réseau est actif sur tous les continents. L'opportunité, offerte par Internet, de partager des connaissances sans délai, depuis n'importe quel pays a stimulé le développement de systèmes d'informations à visée interopérables. Il a donc fallu établir des protocoles d'écritures des connaissances, interprétables partout dans le

monde. Parmi ces institutions, il convient de distinguer les institutions traditionnelles des institutions nouvelles. Les premières sont des productrices de données ou des garantes de la qualité de la production d'informations, depuis le milieu du XXe siècle. Les secondes sont apparues dans le courant des années 2000 en même temps que le réseau Internet et le Web. Pour la première catégorie, nous pouvons citer, notamment, les Universités, l'Organisation Internationale de Normalisation, les instituts nationaux de cartographie comme l'IGN en France ou l'Ordnance survey au Royaume-Uni ou encore les services étatiques responsables, lorsqu'une décision législative le requiert, de l'ouverture des données publiques. La seconde catégorie réunit des institutions fondées afin de répondre aux nouveaux enjeux apportés par la création d'Internet et du World Wide Web (WWW ou Web). Nous retenons le World Wide Web Consortium (W3C)3, la Free Software Foundation, l'Open Source Intiative, la Fondation Wikimédia ou encore la Fondation OpenStreetMap.

Le troisième type d'acteur que nous avons repéré au sein de ce système de production regroupe les entreprises, les auto-entrepreneurs, ainsi que tous les diffuseurs qui mobilisent de l'information géographique libre pour en tirer un profit économique. De nouveau, nous divisons ces types d'acteur en deux catégories. La première catégorie compte les fournisseurs de données commerciales et de services basés sur l'information géographique dont l'ouverture a précédé le réseau GPS et Internet. Parmi eux, nous retenons par exemple les entreprises Michelin, Navteq et ESRI et les instituts nationaux de cartographie, lorsque leurs objectifs sont commerciaux. La seconde catégorie comprend des acteurs commerciaux apparus principalement au début des années 2000 comme les entreprises Google, Telenav, TomTom ou encore Mapbox.

Une fois distingués ces trois types d'acteur, nous tentons de croiser leurs perspectives. Ainsi, les acteurs institutionnels et commerciaux se rejoignent-ils parfois comme autorités méthodologiques ayant vocation à favoriser la standardisation des protocoles

3 Le World Wide Web Consortium (W3C) se décrit comme une communauté internationale au sein de laquelle les membres (des organisations et entreprises telles qu'AirbnB, Amazon, Apple Inc., Google Inc., l'Institut National de Recherche en Informatique et en Automatique (INRIA), l'Open Geospatial Consortium (OGC), l'Ordnance Survey et la Wikimedia Foundation), du personnel à temps plein et le public travaillent ensemble à l'élaboration de normes du Web. Dirigé par l'inventeur du Web, Tim Berners-Lee, le W3C s'est donné pour mission de « permettre au Web de réaliser tout son potentiel ». Le W3C tient un rôle primordial dans notre travail en tant que centre névralgique des recommandations à suivre en faveur d'une interopérabilité entre les techniques de partage des ressources en ligne. L'institution à but non lucratif ne tient pas un rôle prescriptif mais de conseil. Elle détient une influence notable du fait de ses membres fondateurs que sont Tim Berners Lee, le MIT et le CERN. En outre, le W3C est administré par quatre institutions de recherche en informatique : le MIT aux États-Unis, l'ERCIM (European Research Consortium for Informatics and Mathematics) en Europe, l'université Keio au Japon et l'université Beihang en Chine.

de communication et des formats de données. Ce processus de standardisation au service de l'accroissement sensible des échanges d'informations est fortement associé au réseau Internet. À la pointe de la technique, de nouveaux groupes sociaux portent le réseau, le diffusent et l'augmentent techniquement. Survient alors un second processus, également lié à l'internationalisation des échanges, qui bouleverse le processus ordinaire de la production d'informations : le mouvement du Libre. Le Libre est un courant d'idées, au fondement de certains regroupements sociaux, physiques et en ligne, associés aux techniques d'Internet. Ce mouvement est porté par une partie de chacun des trois types d'acteur que nous distinguons. Nous nous intéresserons à la partie fortement active en ligne de ces groupes composés de contributeurs. Si ces groupes constituent souvent des associations internationales, comme la Fondation OpenStreetMap, leurs organisations internes sont très lâches et accordent une place importante à des acteurs industriels parfois moteurs dans le développement de productions contributives. On se demandera dans quelle mesure ces acteurs peuvent s'intégrer aux fonctionnements et aux principes du Libre ou s'ils se rapprochent davantage de la figure des « passagers clandestins » de la théorie des biens communs. Nous verrons que cette question est fondamentale pour envisager la pérennité du système du Libre.

Le tournant d'Internet a provoqué un changement radical dans les pratiques des institutions, contraintes de répondre aux attentes de ce nouveau média. Internet accélère le rythme des échanges des informations ; c'est même son rôle majeur. Des institutions traditionnelles sont alors soumises à la pression : il s'agit pour elles de réduire toujours davantage leurs délais pour publier leurs travaux, au risque de réduire du même coup la qualité des produits finis. Cependant, la contrainte des délais qui pèse sur les institutions est seulement l'aspect le plus visible des changements de méthodes imposés par Internet. En effet, la publication de documents n'est désormais jamais pérenne car toujours perfectible et régulièrement mise à jour sans nouveaux coûts. L'intégration de cette révolution dans les méthodes et les usages caractérise les principes des nouvelles institutions. Il en découle que la logique de fonctionnement de ces nouvelles institutions est forcément différente de celle des institutions traditionnelles.

Considérant ces bouleversements méthodologiques et communicationnel, facteurs de l'émergence du projet contributif OSM, le principal questionnement que nous souhaitons soulever par cette recherche et qui intéresse directement le monde

économique, est le suivant : **est-il possible de confectionner des cartes à partir d'une base de données contributive, libre et diffusée sous une licence ouverte, qui respectent le lien de confiance établi de longue date entre les lecteurs et les producteurs de produits basés sur de l'information géographique ?**

Entre mars 2016 et mars 2019, au sein de l'entreprise Michelin, notre activité de recherche personnelle sur la production de données libres, s'est mêlée à une activité de recherche et de développement à but commercial, répondant alors concrètement à la question ci-dessus. Afin de satisfaire l'entreprise d'une part et d'autre part, pour mettre en œuvre la question de recherche ainsi soulevée, nous avons mis en place deux dispositifs de production de cartes : une série de cartes touristiques papier et une carte numérique plus généraliste et routière. La réalisation de ces dispositifs s'est appuyée sur le cadre méthodologique en construction, du domaine de l'information géographique. Ce cadre méthodologique a accompagné la production des principales sources d'informations géographiques qui font aujourd'hui autorité. Cependant, la construction progressive des systèmes d'informations s'effectue en même temps que les concepts s'affinent. Ainsi, au cours des années 2000, il est apparu qu'une remise en question des systèmes de production apporterait un renouvellement et une valorisation de l'accumulation des connaissances déjà acquises. **L'interopérabilité et la standardisation des formats se sont alors imposées comme des problématiques centrales de la thèse.**

Une de nos premières exigences à propos d'OSM a été de définir la terminologie la plus adéquate. Pour désigner le résultat de la production de la communauté OSM, nous avons, dans un premier temps, envisagé la formule « information géographique volontaire et libre ». Le terme « volontaire » (inspiré du *Volunteered Geographic Information* proposé en 2007 par Michael Goodchild) contenu dans cette formule, permet en effet de distinguer la production OSM des informations involontaires, c'est-à-dire produites à l'insu des internautes. Le terme « libre », quant à lui, permet de souligner le cadre juridique de diffusion qui assure la légalité de l'exploitation de la production par tous et même dans un contexte commercial. **Nous choisissons de ne conserver uniquement l'adjectif « libre » pour qualifier les productions OSM avec l'appellation « informations géographiques libres ».** En effet, le Libre est un système de pensée qui ne se réduit pas à la diffusion ouverte à tous. Il permet ainsi de renvoyer

assez clairement à un nouveau mode de production, ouvert à tout producteur et construit sur l'accumulation progressive des connaissances et des techniques.

Une autre exigence pour une approche rigoureuse d'OSM s'est également imposée : il convient de distinguer les moyens et les enjeux d'OSM lors sa création et, plus récemment, dans les toutes dernières années. En effet, à la création du projet, l'approche *open source* en SIG n'était pas aussi élaborée qu'en 2019, simplement parce que certains outils utiles à l'exploitation d'OSM n'existaient pas. Nominatim, Overpass, OSRM, Imposm sont des outils spécifiques, non fondamentaux, à l'exploitation de la donnée OSM. En quelques années seulement, leur invention a apporté davantage de flexibilité dans la manipulation des bases de données. En 2004, date de création d'OSM, le comité technique Information géographique/Géomatique (ISO/TC 211) de l'Organisation Internationale de Normalisation (ISO) et l'Open Geospatial Consortium (OGC)4 existaient déjà depuis dix ans. Pourtant ce n'est qu'à partir des années 2000 que l'organisation de l'écosystème *open source* du SIG s'est développée. Dans cette évolution technique, certaines dates sont décisives, notamment la première publication en 2002 du logiciel libre SIG QGIS, construit sur la bibliothèque libre GDAL publiée en 2000 et la fondation en 2006 de l'ONG Open Source Geospatial Foundation 5(OSGeo). La standardisation des formats de données a été nécessaire pour le développement de techniques *open source* qui reposent, dans leur principe, sur le partage des techniques et sur l'interopérabilité.

La thèse s'organise en quatre chapitres qui explorent le système de production de la base de données OSM, dans le but de produire une cartographie opérationnelle et de qualité destinée à la commercialisation.

Dans le premier chapitre**, nous évaluons l'ampleur des transformations dans laquelle se trouve engagée la pratique de la cartographie contemporaine, suite à**

4 L'Open Geospatial Consortium (OGC) est une organisation internationale à but non lucratif, fondée en 1994, dans le but de fixer des standards parmi les protocoles et les formats de documents relatifs à l'information géographique. L'organisation garantit la définition des standards basés sur le consensus. Parmi ses membres, dans différentes commissions, on trouve : European Space Agency (ESA), Ordnance Survey, US Geological Survey (USGS), ESRI, Google, le Bureau de Recherches Géologiques et Minières (BRGM) ou encore l'Institut national de l'information géographique et forestière (IGN).

5 L'Open Source Geospatial Foundation (OSGeo) s'est donnée pour mission de favoriser l'adoption mondiale de technologies spatiales ouvertes par un développement participatif et communautaire. L'OSGeo et l'OGC sont partenaires depuis 2009 afin d'allier leurs missions. Le premier se concentre sur la conception d'outils et la production de données tandis que le second vise à encadrer la définition de normes d'écriture des données géographiques.

l'expansion de la discipline vers le contributif en ligne. Ces évolutions concernent autant les normes de production que de diffusion. Traditionnellement, la fiabilité des travaux est garantie par un ensemble de mécanismes culturels et institutionnels. Depuis la création des SIG, puis la diffusion d'Internet et l'invention du Web, la discipline de la cartographie connaît de grandes évolutions méthodologiques et sociétales dans le processus de collecte de données géographiques. L'information géographique volontaire est alors le produit de nouveaux systèmes de production, issus de ces trois bouleversements dans la pratique de la gestion d'informations.

Le deuxième chapitre développe **l'analyse des formes nouvelles de production et de gestion des informations identifiées** dans le premier chapitre **à partir du cas d'OSM. L'évaluation de la qualité et de la fiabilité de la base de données OSM** constitue un élément charnière entre la production des données et la production cartographique. La relecture et la correction manuelle sont en principe assurées par les diffuseurs d'informations, qu'ils soient producteurs de bases de données ou éditeurs de livres ou de revues (Stvilia et *al.*, 2007). Ces mécanismes de garanties de fiabilité ne correspondent pas aux productions issues du Libre. L'exigence de l'évaluation de la qualité de l'information s'est accentuée avec l'apparition des informations originales que sont les informations volontaires. Cette qualité est elle-même relative aux besoins du cartographe. La précision, l'exactitude et la fiabilité d'une information géographique sont des critères permettant d'évaluer la qualité des données. Ainsi, selon le besoin du cartographe, une source ou un ensemble de sources seront préférés pour leur système de référence, pour les garanties commerciales qui accompagnent leurs usages ou parfois pour l'accessibilité à l'information. La première section de ce deuxième chapitre expose l'évolution des travaux scientifiques relatifs à l'évaluation de la qualité des données OSM. Les méthodes classiques d'évaluation de l'exactitude de position, l'exactitude attributaire, la cohérence logique et l'exhaustivité, sont des paramètres construits selon le principe qu'un jeu de données est homogène dans sa modélisation. Cette logique ne peut s'appliquer directement à OSM. Il est donc nécessaire de compléter les méthodes d'évaluation des données s'agissant d'OSM. L'évaluation de la qualité des données repose également sur la compatibilité d'une source de données diffusée avec les besoins, les connaissances et les moyens techniques de l'utilisateur. La deuxième section de ce chapitre s'attache à montrer que le mode de production des données OSM repose sur le processus scientifique de rectification. La méthode de rectification choisie

est elle-même issue des concepts du système Libre selon lesquels la création itérative et ouverte mène au raffinement de l'objet produit. Le caractère itératif d'une production libre permet la rectification tandis que le caractère ouvert soutient une rectification par le plus grand nombre.

Après avoir établi les principes d'évaluation et de fiabilisation des données, le troisième chapitre montre **selon quelles modalités la production de cartes peut s'opérer à partir de la base de données OSM**. Nous nous appuyons, dans un premier temps, sur la réalisation de deux chaînes de production de cartes réalisées chez Michelin. La première chaîne de production sert à la publication d'une carte mondiale et multi-échelles en ligne. La seconde chaîne permet la publication de cartes et de plans papiers touristiques. La suite du chapitre, présente les différents usages cartographiques et les positionnements commerciaux choisis par certaines entreprises internationales qui exploitent les données OSM. Ce chapitre se termine par l'exposé des étapes techniques qui permettent la réalisation effective de cartes papiers et numériques.

Au-delà de l'articulation méthodologique entre la production des données et la production cartographique, le quatrième chapitre aborde **l'articulation des nouvelles opportunités techniques développées via Internet qui motivent la constitution de communautés de bénévoles et qui mettent davantage en relation entreprises et internautes**. Internet est un contexte de production où s'expriment des choix idéologiques et politiques divers. La production d'informations géographiques volontaires, par des bénévoles qui s'organisent en une communauté, se réalise en même temps que le développement de la captation d'informations géographiques involontairement créées par des internautes. Nous nous plaçons dans une perspective exploratoire à la recherche d'indices qui expliqueraient la croissance simultanée de ces productions contraires au cours des quinze dernières années. Nous proposons, au terme de ce chapitre, différentes perspectives qui peuvent être envisagées pour la production et les usages des données OSM.

PARTIE I : Informations géographiques à l'ère du numérique

La cartographie est en cours d'évolution. En quelques années, les géographes et les cartographes ont été les témoins de bouleversements profonds de leurs pratiques. Il est sans doute urgent de réfléchir à cette mutation pour éventuellement l'orienter. Comme le signale Thierry Joliveau en 2010, « Sans doute est-il temps de combler le hiatus ancien qui existe entre la pensée géographique et les outils techniques de traitement de l'information. Ceux-ci ne sont-ils pas, aussi, des outils de la pensée » ?

Nous interrogeons dans cette partie le paradigme lié à la conception cartographique et posons la question de la réalisation cartographique menée à partir d'une base de données contributive, libre (dans sa modélisation et dans sa production), répondant aux besoins qualitatifs du cartographe et diffusée sous une licence ouverte, tout en respectant la convention tacite entre le lecteur de la carte et son producteur d'après laquelle la sélection éditoriale n'est pas dictée par une déficience d'informations.

Notre réflexion se concentre sur les dispositifs de production de cartes qui s'intègrent dans l'exploitation des nouvelles techniques d'Internet ou des nouvelles ressources documentaires, tout en respectant le cadre méthodologique du domaine de l'information géographique. Ces méthodes connaissent une évolution rapide depuis une quarantaine d'années et pourtant des données de référence sont toujours publiées grâce à elles. Cette construction progressive des systèmes d'informations est réalisée en même temps que les concepts se précisent. Au cours des années 2000, il est apparu qu'une remise en question des systèmes de production apporterait un renouvellement et une valorisation de l'accumulation des connaissances déjà acquises. L'interopérabilité et la standardisation des formats s'imposent donc comme des problématiques centrales de la production d'informations. Nous décrirons les procédés méthodologiques mis en place par des communautés d'internautes pour construire des bases de données géographiques qui, tout en sortant des modèles techniques et méthodologiques traditionnels, s'appuient sur des normes. Ces données modélisées sont utilisables en cartographie après de nouvelles transformations propices à la compréhension du message visuel.

Depuis la vulgarisation de l'outil informatique et jusqu'à la fin des années 2000, la carte sur Internet a été un support majeur de visualisation de l'information géographique. Le fond de carte, carte de base servant de localisation géographique au phénomène représenté, peut être complété par d'autres informations thématiques accessibles, manipulables par tous et produites par des experts (Lambert et Zanin, 2012). Dès 2004, Google s'est imposé sur Internet comme un leader en proposant un service en ligne de cartes du monde. A ces avancées techniques d'Internet vers l'interactif et le contributif (Web 2.0) ont succédé les progrès technologiques de production d'information géographique sur le Web (Géoweb). On désigne sous le nom de Géoweb[6] (contraction de Geospatial Web) l'association d'une information géoréférencée et d'Internet, au sens d'un media de partage de ces informations (Palsky, 2013). Le Géoweb constitue ainsi un ensemble de techniques et d'outils permettant aux utilisateurs de créer leurs cartes à partir d'un navigateur Web, en combinant leurs données thématiques (comme des photos géolocalisées via le service Flickr) avec les données provenant d'autres sources comme Google Maps. De cette association (dit *mashup* cartographique) sont issues des informations composites (Lambert et Zanin, 2012). Dans un premier temps, le *mashup* cartographique a généré de nouveaux contenus en combinant l'information. Le plus souvent, la carte est le produit d'un travail individuel et les choix cartographiques sont souvent limités à la localisation de points. Dans un second temps, de nouveaux cadres de production ont offert des possibilités de création directe et collective d'informations géographiques.

La communauté d'internautes concernée par le Géoweb, aujourd'hui transformé en un Web de production de données géographiques, s'est élargie. Cette communauté façonne et use des fonctionnalités numériques récemment devenues plus accessibles, du fait des possibilités de personnalisation de la production sur Internet. Dans ce contexte, les internautes peuvent être des acteurs volontaires de la création de l'information géographique. Il s'agit là d'une révolution dans le mode de production de l'information géographique qui ne concerne aujourd'hui qu'une minorité de personnes. Cette nouvelle possibilité de diffusion libre de données et le mouvement Libre qui

6 Défini par Thierry Joliveau comme « une organisation de l'information d'Internet à travers un géoréférencement direct ou indirect sur la surface terrestre de tout contenu informationnel » (Joliveau, 2011). Il associe l'émergence de cette forme de structuration des informations aux services de recherches de ressources en ligne proposés par l'entreprise Google. Google Inc. mobilise son service Google Maps pour localiser et représenter par des épingles les résultats aux requêtes des internautes utilisateurs de son moteur de recherche.

l'accompagne, encourage la participation du plus grand nombre. OpenStreetMap (OSM) est issue de cette extension de la pratique contributive en communauté.

Le projet OSM s'est développé grâce à ce contexte technologique des années 2000. Le processus de production d'informations géographiques volontaires a été rendu possible notamment par l'accès gratuit aux signaux GPS. En outre, grâce au Web, les communications simplifiées entre individus, où qu'ils soient dans le monde, ont permis le développement synchronisé et rapide de la carte mondiale OSM. Des « citoyens » volontaires, appelés « contributeurs » se sont engagés dans la conception, la création et l'amélioration d'une base de données géographique commune. Cette organisation communautaire en ligne est permise par une infrastructure *open source*[7]. Le fonctionnement autour d'outils dont le code source est ouvert signifie un renouvellement fréquent des moyens techniques et de la folksonomie structurée[8] de la base OSM. Cette instabilité de l'environnement explique en partie le succès actuel du projet mais exige l'investissement de nouveaux contributeurs et une veille constante des évolutions par les contributeurs initiés. En une décennie, le projet a connu un essor considérable dans le monde puisque depuis sa création en 2004, un million de contributeurs l'ont rejoint. L'association OpenStreetMap vise à produire et à fournir des données cartographiques libres à quiconque souhaite les utiliser. Aujourd'hui, OpenStreetMap, qui use des nouvelles fonctionnalités de contribution offertes par Internet, réunit aisément des contributeurs qui participent à la construction d'une même base de données, en y apportant des connaissances empiriques sur leur environnement. Les contributeurs étoffent les descriptions relatives à leur terrain pour garantir une information à jour. Cette pratique cartographique qui paraît se développer, en dehors de tout cadre (méthodologique et institutionnel), est destinée à durer car elle est portée par le développement de produits en direction du grand public.

7 Nous précisons dans la sous-partie I.3. Les droits d'utilisation des données spatiales, la distinction entre l'appellation *open source* et l'appellation Libre. L'*open source* est « un label dissident » (Xifaras, 2010) du mouvement du Libre apparu en 1998 avec l'accroissement des intérêts économiques du développement des logiciels, d'une part et de la pratique du partage des codes sources, d'autre part. Des projets de développement logiciel ou de ressources documentaires en ligne peuvent être cohérents avec les principes techniques du Libre, c'est-à-dire la mise à disposition à tous les internautes des ressources mais en contraction avec les principes moraux du mouvement du Libre, qui consistent en la défense de la vie privée à protéger en priorité face aux attentes de rentabilité (Xifaras, 2010).

8 Le terme désigne à la fois une pratique de production et son résultat. La pratique consiste en l'indexation par les internautes des ressources en ligne. Les ressources sont qualifiées par des étiquettes (*tags*) affectées par les internautes. La définition est précisée est la pratique est confrontée à nos questionnements dans la section I.1.2. pour l'interopérabilité des bases de données géographiques.

Nous partons dans cette partie du constat de l'hétérogénéité inédite et déroutante de la manière de relever l'information lorsque cette information est renseignée librement. En effet, les méthodes de relevés de terrain ou de descriptions d'information sur une carte varient fortement selon les contributeurs qui participent à un projet collectif. Nous formulons alors l'hypothèse que cette difficulté réelle ne constitue pas un obstacle insurmontable à la transmission réussie de l'information par l'image cartographique. Le message de la carte est compris lorsqu'un compromis est trouvé entre la liberté des méthodes de relevés et l'observation de règles de traduction graphique d'informations géographiques (sémiologie graphique).

Le projet OpenStreetMap repose également la question de la confiance envers le producteur de la donnée. Depuis plus d'un demi-siècle, les producteurs tels que les instituts nationaux ou les producteurs privés de référence comme Google et TomTom ont su asseoir leur autorité dans le domaine de la cartographie par la démonstration de l'efficacité pratique de leurs cartes. Cette efficacité repose sur les moyens mis en œuvre et sur les compétences techniques développées par ces producteurs. Le contexte contributif où tout individu peut renseigner une information géographique ne s'appuie que partiellement sur ces compétences et repose sur une organisation de production radicalement différente. Peut-on obtenir des informations qui mèneront à des cartes d'efficacité équivalente en usant de moyens de collecte différents et de compétences diversifiées ? Faut-il changer de modes de représentation afin de bénéficier des informations ainsi produites ? Les habitudes des utilisateurs des cartes sont-elles à bousculer ? Notre approche ne consiste pas à évaluer ni à statuer de manière définitive sur la validité du projet d'une base de données géographique libre, mais vise à mieux comprendre quels sont les enjeux actuels liés à la construction et à la diffusion des savoirs.

Chapitre 1 : Changements de paradigme en cartographie

De l'ordinateur-calculateur au tournant du numérique

En cartographie, deux changements de paradigme ont eu lieu : l'un dans les années 1960, suite à la création du système d'information géographique qui représente un tournant méthodologique et technologique majeur et l'autre dans les années 1990, en réponse au procédé d'ouverture des codes sources et des données géographiques qui constituent un bouleversement à la fois épistémologique et industriel.

Si les années 1960 marquent, avec l'apparition de l'ordinateur, un tournant dans le rapport à l'information géographique, une longue période transitoire avait, en fait, commencé bien avant. On prend souvent comme repère l'exemple de la carte du choléra à Londres, publiée en 1854 par John Snow. En superposant sur une même carte un symbole pour chaque personne décédée des suites du choléra et d'autres symboles caractérisant leur contexte environnemental, John Snow révèle une relation entre ces deux couches d'informations. La méthode est remarquable car Snow conçoit ainsi deux couches de données cohérentes thématiquement, topographiquement et topologiquement. Par cette mise en relation d'informations spatialisées, John Snow produit de l'information géographique dont l'élément caractéristique est la reconnaissance de l'organisation systémique d'une occupation dans l'espace. Chaque information doit être associée à son contexte environnant, afin de lui donner un sens géographique. Si cette conception, en couches de données, est aujourd'hui pratique courante, elle est encore minoritaire dans la pratique de la géographie du début du XXe siècle. Plus largement, au début du XXe, la représentation visuelle de l'information géographique en vue de communiquer des connaissances ou de poser un raisonnement n'est que modérément pratiquée. En fait, cette lente progression du changement de paradigme a duré jusqu'aux années 1990. Cependant, cette intégration n'est, au début des années 1960, qu'un changement d'outil et non pas de méthode. Au début des années 1960, l'intégration de la problématique de l'analyse spatiale dans des recherches fondamentales – exploratoires ou appliquées – est pratiquée manuellement par les chercheurs. Ce changement s'intègre dans un processus plus large : les débuts de l'analyse de données (Tukey, 1961) et de l'informatisation des méthodes de traitement. Les méthodes retenues ne permettent pas alors de manipuler une quantité d'informations suffisante pour répondre aux besoins de l'analyse spatiale, sauf aux

termes de travaux de collecte de longue haleine. Déjà, cependant, à la fin des années 1960, des chercheurs comme Roger Tomlinson (1968) et Howard Fisher (1983) contribuent à la transformation de ces concepts et de ces méthodes d'analyse spatiale manuelles en méthodes informatiques. C'est dans ce cadre que les récents principes méthodologiques et conceptuels des systèmes d'informations sont mobilisés par les géographes.

La diffusion des systèmes d'informations géographiques, survenue entre 1980 et 1990, années de la démocratisation des outils informatiques, est consécutive à la révolution informatique. Dans le domaine géographique, cette évolution a rendu accessible à la recherche et à certaines entreprises ou à certaines institutions les moyens de construire des systèmes d'informations. En cartographie, l'écriture d'informations géographiques conçues pour être intégrées en bases de données est passée d'une pratique minoritaire à une pratique dominante. On peut ainsi évoquer sans excès un changement de paradigme, même si une longue période transitoire a été observée, qui a vu coexister puis s'entremêler la pratique du dessin manuel et le traitement automatique des données. Ainsi, entre les années 1960 et les années 1990, on a pratiqué une cartographie informatisée sans réellement profiter de l'apport du nouvel outil informatique et sans modifier notablement les pratiques anciennes des cartes dessinées à la main. L'opportunité d'un tri automatique assisté par l'ordinateur et d'une hiérarchisation systématique n'était que rarement envisagée et c'était sur l'expérience du cartographe que reposait l'harmonie graphique mais également la cohérence des objets sélectionnés. Dans ce cas, l'apport de l'informatique a probablement consisté en une économie de temps de travail et pas encore en une uniformisation des rendus.

Depuis les années 1990, le paradigme s'inverse et le numérique impose sa logique. La création de cursus universitaires dédiés, de programmes de recherche, de comités scientifiques et techniques internationaux, de logiciels commerciaux et de bases de données géographiques institutionnelles ou commerciales en attestent. L'apparition des systèmes d'informations est un premier bouleversement dans la pratique de la cartographie, suite aux nouvelles opportunités d'automatisation et de pérennisation des informations. Il est suivi par une véritable révolution dans la conception même des cartes produites.

Le second changement de paradigme, plus récent, présente une dimension politique. Dans le cadre du courant dit Libre9, initié au début des années 1980, les technologies soutenant des systèmes d'informations géographiques deviennent plus accessibles. Entre 1999 et 2004, le protocole Web Map Service10 (WMS, 1999) de partage d'images géolocalisées était proposé par l'Open Geospatial Consortium ; Bill Clinton, président des États-Unis d'Amérique, ouvre au grand public l'accès au réseau GPS (2000) et les premières versions de l'extension à PostgreSQL11, PostGIS12 et Qgis (logiciel de SIG libre) sont publiées. Ces événements sont représentatifs d'un ensemble d'éléments techniques spécialisés (Qgis et PostGIS), sociaux (la construction d'institutions responsables de garantir ces publications) et politiques (le mouvement d'ouverture des techniques et des données) qui participent d'un changement de paradigme fort, relatif à l'information géographique. En outre, à ces mêmes dates, on assiste à l'émergence de l'encyclopédie en ligne Wikipédia, créée et enrichie par des internautes anonymes, mais aussi de la carte en ligne mondiale et multi-échelle de Google (2005), du Géoportail de l'IGN (2006) et du projet de base de données contributive OpenStreetMap. Le courant du Libre trouve un espace d'expression au travers du réseau Internet, d'abord par la publication de connaissances statiques, c'est le web 1.0. Puis, le web 2.0 en appelle à la contribution des internautes afin de créer des contenus de connaissances évolutifs et des outils.

Ces deux changements de paradigme aboutissent à la généralisation des systèmes d'informations, dont une grande partie des technologies sont déjà ouvertes. Désormais, l'information géographique s'écrit sous la forme de bases de données. La part de la description traditionnelle des paysages tend progressivement à se réduire au regard de la grande quantité d'informations que les systèmes d'informations géographiques permettent de créer. La publication des connaissances dans des bases de données s'impose aujourd'hui dans le contexte de publication libre qu'est Internet. Elle

[9] Le courant du Libre se présente, au travers de personnalités et d'institutions, comme un mouvement social dont les pratiques de diffusion des connaissances, placées en libre accès et de repartage des modifications constituent les fondamentaux.

[10] Le Web Map Service est un protocole de partage d'images géoréférencées. Ce protocole est un standard garanti par l'Open Geospatial Consortium depuis 1999. Il permet de visualiser des cartes sous la forme de tuiles raster en ligne ou par le moyen d'un SIG utilisé sur un ordinateur connecté à Internet.

[11] PostgreSQL est un système de gestion de base de données (SGBD) relationnelle publié sous licence libre et diffusé depuis 1997.

[12] PostGIS est une extension au SGBD PostgreSQL. Elle permet la manipulation d'objets géographiques vectoriels dans le respect des standards de l'OGC. Publiée depuis 2005, cette extension permet de manipuler des bases de données géographiques depuis des SIG.

bouscule les activités industrielles dont le modèle économique repose encore sur la propriété intellectuelle d'informations publiées sous une forme cartographique.

De la numérisation aux méthodes numériques

L'information géographique sur Internet, est peu présente dans le web 1.0 mais elle est centrale dans le web 2.0. Le web 2.0 a permis l'accès aux informations stockées dans les bases de données situées sur un serveur et même d'y ajouter des informations supplémentaires. Ces systèmes d'interactions avec des bases de données, accessibles par le réseau Internet, permettent à tous de produire de l'information ouverte, notamment de l'information spatialisée. Ce Web (2.0) de production d'informations spatialisées a émergé grâce à l'intégration des moyens comme le SIG GRASS13, l'extension PostGIS du Système de Gestion de Base de Données (SGBD) PostgreSQL et le protocole WMS dont les technologies ont été intégrées progressivement à des environnements de développement d'application à référence spatiale comme MapServer (1994). Dans le cadre du Géoweb, il est devenu possible pour le grand public de localiser une information en ligne. Ainsi, les bases de données géographiques éditables en ligne par tous sont-elles nées du croisement de la volonté d'artistes et de développeurs de créer des biens communs de la connaissance, protégés du risque d'appropriation et des moyens techniques que sont les SIG. Cet objectif de construction de biens communs de la connaissance se réalise depuis une dizaine d'années au travers du web 3.0. qui est un Web plus interopérable et davantage « hyperlié »[14]. Le web 3.0 présente en outre l'avantage de s'appuyer sur les bases de données géographiques déjà conçues dans le cadre du Géoweb. Le web 2.0 n'a pas rendu caduc le web 1.0 et les éléments d'un web 3.0 – un web à venir dont les spécificités consistent en une montée en complexité des réseaux physiques et sémantiques – coexistent déjà avec les deux niveaux d'interactivité précédents. Le faible niveau d'interactivité des débuts du web 2.0 permet seulement de créer de l'information spatialisée grâce au mashup. Le développement du Géoweb et des *wikis*[15], ouvrent le cadre de production de l'information réellement géographique par l'apparition de réflexions communes. Par

[13] Logiciel de traitement de données géographiques publié en 1983. GRASS est un SIG publié sous licence libre dont de nombreuses fonctionnalités sont intégrées au SIG libre QGIS.

[14] Nous faisons référence aux hyperliens qui participent à la connexion entre les ressources sur le Web

[15] « Un *wiki* est une application web qui permet la création, la modification et l'illustration collaboratives de pages à l'intérieur d'un site web. Il utilise un langage de balisage et son contenu est modifiable au moyen d'un navigateur web. C'est un logiciel de gestion de contenu, dont la structure implicite est minimale, tandis que la structure explicite émerge en fonction des besoins des usagers. » (Wiki., 2019, dans Wikipédia, l'encyclopédie libre)

exemple, les couches créées dans la base de données OpenStreetMap sont conçues dans le cadre d'un modèle conceptuel de la réalité, co-écrit par des contributeurs volontaires. Avec OpenStreetMap et Wikipédia, l'interactivité permise est au maximum de ce que permet actuellement, en 2019, le web 2.0.. Par ailleurs, dans ce travail, nous examinons le processus de co-conception d'informations géographiques libres en « suivant le médium » (Severo, 2017) qu'est le *wiki* du projet OSM, au travers de ses évolutions chronologiques mais également dans ses spécificités nationales.

L'accroissement des capacités de stockage, des puissances de calculs et d'un ensemble de méthodes d'analyse des données soutient la constitution du gisement de connaissances chaotiques qu'est devenu le Web. Dans les années 1990, à ce web devenu trop dense à parcourir sans outils adaptés, Tim Berners Lee propose comme alternative l'idée d'un Web, dont les documents ne seraient pas uniquement référencés et auto-référencés mais dont les références seraient porteuses de sens. Ces liens – comparables aux hyperliens des pages HTML – qualifiés, organisés et hiérarchisés, seraient aisément décodés aussi par un être humain ou par une machine. Afin de soutenir cet objectif d'un web davantage lié et dont les liens seraient porteurs de sens, l'idée du Web sémantique commence à se répandre. Le W3C en propose des règles d'implémentation, structurées selon les langages RDF (Resource Description Framework) et OWL (Web Ontology Language). Ces formalismes descriptifs des ressources d'Internet sont à l'origine de la création de nouvelles bases de connaissances. Ces bases sont des espaces de stockage de graphes représentant les savoirs et leurs relations qui sont enregistrés sur Internet mais qui sont également disponibles sous leur forme matérielle, comme des livres ou des films. Dans ces bases de connaissances, les objets stockés sont associés à d'autres objets de la même base de connaissances mais également à des documents du web (page HTML) ou à d'autres objets provenant d'autres bases de connaissances et de bases de données relationnelles. Des communautés, constituées notamment de développeurs et de contributeurs, construisent l'architecture informatique du web 2.0 pour en faire l'interconnexion universelle de sources de données à l'échelle du monde (Abiteboul et *al.*, 2014). Les bases de connaissances en question présentent un niveau de généralisation supérieur à celui de toutes les bases de données déjà créées en ligne. Les connaissances géographiques qui sont stockées selon la logique des bases de connaissances peuvent être interrogées et confrontées à d'autres connaissances

connexes appartenant à n'importe quel domaine du savoir (littérature, histoire, biologie) afin de déduire de l'information géographique nouvelle.

Ainsi, la cartographie est une discipline fortement transformée par les évolutions technologiques de communication. En outre, ces transformations se poursuivent rapidement. L'information géographique est majoritairement produite selon l'écriture en bases de données. Cette forme de production suit un cadre conceptuel et un cadre méthodologique encore en développement. C'est ce que nous expliquerons en première section de ce chapitre (I.1. Formalisation informatique de l'information géographique).

Le second paradigme que nous avons décrit, celui de l'imprégnation des modes de production et de diffusion par les principes du Libre, mène à une nouvelle production d'informations géographiques : les informations géographiques volontaires. Ce type de production, réalisé à l'aide des nouveaux moyens technologiques de mesure et de communication, est conçu en communauté. Nous détaillerons en deuxième section (I.2. L'information géographique volontaire, contributive et libre) les modalités de cette production originale et nous développerons l'exemple d'OpenStreetMap.

De manière générale, en cartographie, les bases de données commerciales ou ouvertes sont conçues pour être diffusées. La gratuité sur Internet ne signifie cependant pas un droit d'usage des données et des cartes publiées sans restriction. Nous étudierons les enjeux et les choix de conception des licences de données en troisième section de chapitre (I.3. Les droits d'utilisation des données spatiales).

1.1. Formalisation informatique de l'information géographique

La construction de systèmes d'informations géographiques rapproche le champ de la géographie des domaines d'activité scientifiques et techniques qui mobilisent les théories informatiques. Ainsi, la documentation, l'archivage, l'indexation ou la modélisation des connaissances, qui ont précédé l'apparition de l'informatique et sa démocratisation, sont des domaines qui organisent désormais leurs connaissances à l'aide de bases de données potentiellement accessibles via Internet.

En 1936, André Cholley écrit dans l'éditorial du premier numéro de la revue *L'information géographique* que le but de la revue est de proposer, au service de l'enseignement, des publications spécialistes assurant le renouvellement des connaissances relatives à la discipline de la géographie (Cholley, 1936). Sous la plume d'André Cholley, l'expression « information géographique » désigne un élément de connaissance obtenu par la pratique de la géographie et évidemment sans rapport avec l'informatique. Les informations géographiques sont déjà conçues en couches thématiques, autonomes les unes par rapport aux autres, mais dépendantes dans leur conception, par leur rapport topographique et topologique. L'appellation « système d'information géographique » (SIG) fait de l'information géographique un objet de réflexions méthodologiques visant à organiser des connaissances géographiques en système. Le terme système désigne la conception d'un objet géographique comme nécessairement reliées à ses voisins topologiques et thématiques qui sur des concepts géographiques adaptés aux besoins informationnels auquel le SIG répond. Ces relations peuvent s'exprimer informatiquement au travers du formalisme des bases de données relationnelles.

Le terme d'information est une appellation très large qui suppose un échange intelligible entre l'émetteur de l'information et son récepteur. Une information est un renseignement communiqué à l'aide de mots, de sons ou d'images. Dans le contexte des technologies numériques, cette information peut être transmise par un message, dont la structure est interprétable par un récepteur. La donnée n'existe que par sa possibilité d'interprétation par un utilisateur. Un jeu de données, publié par un producteur, est nécessairement associé à un élément qualitatif qui permet une juste interprétation de la donnée. Si la donnée en mathématiques, est ce qui est admis, connu et qui sert de base à un raisonnement ou de point de départ pour une recherche, en

informatique, la donnée est plutôt l'information qui peut être interprétée, notamment par sa représentation sous forme conventionnelle, destinée à en faciliter le traitement par ordinateur (Rey et Morvan, 2005). Ainsi, nous considérerons qu'une information géographique est le résultat de l'interprétation, qui fait suite à une observation de terrain, à la lecture de réponses à un questionnaire ou encore à l'analyse d'images photographiques ou satellitaires. Une donnée géographique est donc une information géographique, qui est nécessairement encodée pour correspondre à un ensemble de règles de description des informations, ce que l'on désigne sous la terminologie « modèle de données »

Si, comme Thierry Joliveau, on entend « géographique au sens de la discipline », « la donnée » doit « répondre à une problématique géographique » (Joliveau, 2011). Selon cette même pensée, nous considérerons que l'expression « information géographique » désigne toute organisation de connaissances dont le but est de répondre à des problématiques géographiques. Une base de données géographique est le résultat d'une représentation sélective et méthodique du monde réel. Pour être fonctionnel, le mode de production et la structure d'une base de données doivent être documentés grâce à des métadonnées et ses objets doivent être décrits selon un même cadre conceptuel. , Il est, par exemple, important que la topologie des objets, les uns par rapport aux autres, soit conservée dans la base de données géographique. L'échelle qui guide la généralisation nécessaire des formes et des attributs doit être constante. **Nous considérons ainsi qu'une base de données géographique se définit comme un ensemble d'objets qui, pris indépendamment les uns des autres, ne sont que spatialisés et qui, pris dans leur ensemble, sont le résultat de l'analyse d'un territoire**. Ceci implique qu'une base de données géographique est nécessairement porteuse d'un message et d'une subjectivité car elle est construite en fonction d'une modélisation particulière et d'un objectif spécifique. Pour qualifier plus clairement certains jeux de données, sortis de leur cadre de conception et, de ce fait, difficiles à exploiter dans un objectif d'analyse géographique, nous emploierons le mot « spatial » lorsque les objets sont géo-référencés à la place du mot « géographique ». C'est le cas de certaines sources d'informations en *open data*, parfois diffusées partiellement et dont les métadonnées peuvent manquer. La source est alors partielle et n'est plus utilisable pour une analyse géographique. C'est également le cas des données géographiques captées en ligne et relatives aux internautes. La distinction entre les qualificatifs

« géographique » et « géospatial » repose donc sur le travail méthodologique qui précède la production concrète et sur le caractère exhaustif ou non de la publication d'une base de données.

Plus généralement, comme tout document, une base de données doit être accompagnée de métadonnées. Ces données sur des données ont pour rôle de préciser le contexte, les méthodes et les règles qui ont mené la production du document. Le terme de métadonnées, connoté dans le domaine de l'informatique, désigne une pratique de référencement inhérente à toute forme d'indexation par des notices. Les métadonnées sont fondamentales pour une bonne utilisation de la source d'informations numérique comme de toutes sortes de sources documentaires. Dans le cadre de la construction du web sémantique, les métadonnées ont la fonction de relier les informations entre elles et portent des informations relatives à l'auteur du document, à la date de la création, à la dernière mise à jour, à la nature du document (base de donnée, photographie, trace GPS), et à son contenu (par le moyen de mots-clés). L'indexation par notice respecte un formalisme nécessaire à la recherche de la construction d'un web structuré. Les métadonnées donnent à un document un sens interprétable aussi bien par les humains que par les machines (grâce à la structure de la description). Elles participent donc à un meilleur partage de l'information, et à l'interopérabilité de différents documents produits dans des contextes distincts, lorsqu'elles sont mobilisées dans ce sens.

Le besoin de clarification du contexte de production accompagne l'expansion des publications de bases de données et de cartes en ligne. Une partie de ce contexte de production se compose de la définition du modèle des données de la base. En première sous-partie (I.1.1. De l'observation de la réalité à la base de données géographique) nous détaillons l'évolution, depuis les années 1960 aux années 2000, de la formalisation des modèles de données. L'augmentation du nombre de publications mais également des thématiques, des langues d'écriture et des formes de modélisation donnent accès à des ressources en informations très riches mais incompatibles et potentiellement redondantes. Conscients des limites de la multiplication des sources de données lorsqu'elles sont dénuées de l'explicitation de leurs spécificités, certains producteurs développent des procédés méthodologiques pour obtenir davantage d'interopérabilité entre les sources : c'est ce que nous verrons dans la deuxième sous-partie (I.1.2. Pour l'interopérabilité des bases de données géographiques). Nous présentons également une synthèse du procédé méthodologique, en développement dans le cadre du web

sémantique, permettant à plusieurs sources de données, de partager un cadre conceptuel, d'être exploitées selon différentes méthodes et pouvant conduire à des bases de données interopérables et adaptées à des besoins spécifiques.

1.1.1. De l'observation de la réalité à la base de données géographique : décrire et prescrire une vue de la réalité

Une carte est un modèle d'un espace observable ou imaginé (Board, 1967, Cauvin et *al.* 2007). Elle correspond à une représentation formalisée et épurée du réel (Brunet et *al.*, 1992). La production d'une carte est ponctuée d'un enchaînement d'étapes de transformation, depuis l'information primaire issue des observations de terrain ou de conceptions imaginaires, jusqu'à leur description par une symbologie explicite. Depuis la révolution informatique, la carte ne se dessine plus à la main. Entre 1960 et les années 2010, de nouvelles transformations appliquées aux données attributaires et aux formes des objets géographiques assistées par ordinateur sont formalisées notamment par des chercheurs comme Michael Goodchild, David Maguire, Sylvie Rimbert, Colette Cauvin et J B Harley. Ces chercheurs présentent les transformations nouvellement possibles comme relevant d'un contexte scientifique « exploitant les nouveautés qu'elles soient théoriques, méthodologiques ou technologiques, élargissant ainsi ses possibilités et ses champs » (Cauvin et *al.*, 2007). Ces transformations relèvent de la classification hiérarchique du réel par la définition de ses objets en schémas, par la sélection d'une partie du réel observé et par le niveau de généralisation. Elles constituent, chacune, une étape d'abstraction du réel qui participe de la construction du modèle graphique qu'est la carte. Elles suivent des modèles au sens où leurs procédés de fabrication répondent à un ensemble de règles prescriptives. Les modèles de données constituent des méthodes d'écriture des informations géographiques à l'origine de la carte.

La chaîne de production s'appuie sur un cadre théorique et sur des sources primaires ou secondaires d'informations géographiques. Elle conduit le producteur à l'écriture de données géographiques structurées en une base de données géographique dont la construction présente des aspects classiques mais également des spécificités, liées à son domaine propre. Cette construction s'appuie sur un cadre méthodologique constitué d'un ensemble de schémas et de règles qui caractérise le jeu de données géographiques.

Cette structure méthodologique présente différents niveaux d'abstraction (Balley, 2007) que nous détaillons maintenant.

Un modèle de données est le cadre théorique qui permet de décrire une perception de la réalité. Un modèle géométrique peut le compléter. Par ce modèle, on décide du niveau d'abstraction géométrique des objets. Deux formes d'abstractions géométriques se sont imposées depuis une décennie : l'abstraction par des formes vectorielles ou l'abstraction par carroyage. Les données vectorielles ont pour forme un dessin vectoriel en points, lignes (liste de points) ou polygones (liste de points dont le premier et le dernier sont le même point). Les données raster se présentent sous la forme de carroyages réguliers ou non. Dans le cadre d'une production de données dont le but est de rendre l'information exploitable par un public non producteur, le processus de formalisation informatique de l'information géographique doit être documenté. C'est ce qu'on appelle « le schéma conceptuel de données » qui est une formalisation graphique du processus logique ayant mené à la réalisation de la base de données. Toute base de données doit répondre à un ou plusieurs usages précis. C'est pourquoi chaque base possède un schéma conceptuel spécifique, qui simplifie et explique les sélections des éléments de la réalité à enregistrer. Si la modélisation est un exercice conceptuel, la base de données doit donc être documentée en ce schéma conceptuel de données. Ce schéma comporte un ensemble de règles prescriptives, stipulant la forme élémentaire de chaque objet et ses attributs fondamentaux. Ainsi, si l'on souhaite, par exemple, constituer une base de données des fleuves, on stipulera que l'objet « fleuve » du modèle conceptuel est composé de « tronçons de fleuve » dessinés en une ligne qui suit le chenal du fleuve. Enfin, des règles d'intégrité contraignent les objets de la base de données (Balley, 2007) de manière à n'autoriser que les seuls objets cohérents à cette base. Si on considère qu'un bassin versant est constitué d'un graphe de tronçons de fleuve, chaque tronçon ajouté ou modifié doit être topologiquement cohérent avec le reste du graphe. De ce fait, producteurs et utilisateurs partagent une même classification, dont l'implémentation est exprimée dans le schéma conceptuel. Cette implémentation mène à l'écriture des données en bases de données, à leur utilisation et à leur mise à jour. En géographie, la pratique la plus courante consiste à rapporter chaque type d'objets, identifiés dans le schéma conceptuel, à la table qui lui correspond et à rapporter chaque caractéristique à la colonne qui lui correspond. Ces tables, pour être en relation, doivent partager des attributs communs – des clés – qui permettent de les associer entre elles

afin de constituer d'autres tables ou de croiser les données, si besoin. On emploiera dans ce cas un logiciel de type Système de Gestion de Base de Données (SGBD) relationnel pour enregistrer, modifier et exploiter les données[16].

Aujourd'hui, la donnée est le matériau principal des analyses spatiales et des représentations cartographiques. Ces données sont produites et publiées sous la forme physique de bases de données qui peuvent être utilisées de manière isolée ou en recoupant les informations de différentes bases de données. En géographie, une pratique courante consiste à associer des données géométriques pauvres en informations attributaires et des données attributaires propres à l'analyse souhaitée. En France, le principe de distinction entre les productions de nature attributaire (statistique) et les productions de nature géométrique (topographique) s'est ancré dans la pratique au travers de la création des institutions nationales productrices (Joliveau, 2011). L'Institut national de la statistique et des études économiques (INSEE) est producteur de données sociales et économiques, tandis que l'IGN est en charge du maintien de l'infrastructure des données géographiques nationales. Ces deux acteurs institutionnels garantissent la collecte, l'organisation, le maintien et l'utilisabilité des données qu'ils produisent dans le cadre de missions d'intérêt général fixées par l'État. Dans le cadre du processus d'ouverture des données publiques, l'INSEE et l'IGN sont deux diffuseurs importants de bases de données en ligne. L'exploitation conjointe de ces données publiées par l'INSEE et l'IGN mais également par d'autres acteurs plus modestes, comme certaines villes, rend désormais nécessaire l'usage de référentiels communs. L'INSEE et l'IGN partagent des identifiants communs relatifs aux limites administratives.

À l'échelle européenne, citons plus particulièrement l'organisme de statistiques Eurostat et les projets ESPON qui ont pour visée de constituer « des jeux de données complexes portant sur l'ensemble du territoire européen, et parfois son voisinage, à différentes échelles, dans différents maillages territoriaux ». Les projets ESPON proposent des mesures mêlant différentes sources de données à plusieurs dates. Pour assurer la continuité des études « les jeux de données reposent sur une ou plusieurs nomenclatures du territoire, parfois sur plusieurs niveaux (par exemple NUTS 0, 1, 2, 3), et pour différentes versions (1999, 2003, 2006) » (Bernard et *al.*, 2014). Ces variations des découpages de la nomenclature sont issues des changements territoriaux,

[16] D'autres langages d'implémentation de bases de données peuvent être choisis, comme le XML ou le Java.

suite à des choix politiques nationaux. Ces changements de nomenclature doivent être implémentés dans un modèle afin d'assurer l'interopérabilité des jeux de données à différentes dates. Ainsi, « le modèle supporte plusieurs versions d'une même nomenclature » et l'historicité des modifications administratives peut être retracée (Bernard et *al.*, 2014).

Cette facilitation des croisements de sources par l'interopérabilité entre les jeux de données mais surtout par l'harmonisation des concepts et des définitions guidant la production des données, est en cours de mise en place dans les institutions françaises. La directive européenne INSPIRE de 2007 montre, qu'à l'échelle de l'Union Européenne, les besoins sont importants. En France, jusqu'au milieu des années 2000, les administrations publiques, conservaient encore des bases de données cloisonnées (Joliveau, 2011). La base de données d'une administration représentait davantage un « trésor à protéger et à défendre plus qu'à partager » (Joliveau, 2011). La pratique change car elle est soumise à un processus politique[17] mais également pour des raisons économiques. La mise à disposition en ligne des données a été en effet un moyen facilitateur de réponses à une demande citoyenne. Cependant, c'est d'abord la prise de conscience du gain économique obtenu grâce au partage des charges de maintien des données, parfois redondantes entre les services, et la valorisation de ces données auprès des acteurs économiques et des citoyens qui a motivé le développement de l'interopérabilité entre les producteurs de données (Joliveau, 2011).

L'interopérabilité entre les jeux de données est un objectif accompagné par un processus de standardisation de l'écriture de l'information qui concerne deux sphères de la production de données. Elle peut être appliquée pour des méthodes de collecte et de structures des données. Elle peut également concerner les formats numériques selon lesquels la donnée est écrite. L'Open Geospatial Consortium (OGC) ainsi que l'International Standards Organisation (ISO) diffusent des normes d'écriture de l'information géographique. Si une large partie de leurs préconisations sont approuvées et suivies par les producteurs d'informations géographiques, Brown et *al.* signalent que certains formats se sont diffusés en parallèle. Ils notent l'importance de formats propriétaires comme le *shapefile* d'ESRI[18], finalement ouvert suite à des « pressions »

[17] Développé en 1.3 de ce chapitre.

[18] *Esri shapefile technical description: an esri white paper—july 1998*. Consulté en janvier 2019 sur : url : https://www.esri.com/library/whitepapers/pdfs/shapefile.pdf)

exercées par la communauté d'utilisateurs (Brown et *al.*, 2013). Les spécifications décrivant le mode opératoire et la structure du jeu de données sont deux documents exigés dans un processus standard de partage de données. Ces documents répondent également à des standards décrivant les éléments fondamentaux à préciser comme l'instrument de mesure choisi pour la collecte, la date de la collecte ou encore les contraintes imposées par le schéma conceptuel de données.

L'interopérabilité et la standardisation sont des éléments définissant le niveau d'« utilisabilité d'un jeu de données ». La norme ISO 9241-11 2018 définit l'utilisabilité comme le « degré auquel un système, un produit ou un service peut être utilisé par des utilisateurs spécifiés pour réaliser des objectifs spécifiés avec efficacité, efficience et satisfaction, dans un contexte d'utilisation spécifié »[19]. La problématique de l'utilisabilité devient une priorité pour les producteurs de données comme l'IGN ou l'Ordnance Survey. En effet, « les utilisateurs » ne recouvrent pas seulement le groupe d'experts formés au traitement des données et utilisant des logiciels adaptés à leurs besoins, qui sont les principaux interlocuteurs des producteurs avant la diffusion des données par le moyen d'Internet (Brown et *al.*, 2013). En effet, les instituts dont la mission est de service public ont l'obligation de s'adresser à un public potentiellement aussi divers que la population de leur pays. Différentes initiatives comme Ordnance Survey OpenData, Open Data Paris, data.gouv [France] ou data.gov [États-Unis] publient sous des licences ouvertes des jeux de données auprès de potentiels utilisateurs aux compétences et aux besoins variés. Selon Brown et *al.*, si le développement des systèmes d'informations géographiques a d'abord été concentré sur le développement de fonctionnalités logicielles pour le traitement de la donnée, l'innovation au profit de l'utilisabilité doit être davantage tournée vers la formalisation des principes qui dirigent la constitution d'un jeu de données. Des formats de données incompatibles car ne respectant pas les mêmes standards de structures ou de formats numériques représentent un défi dans une chaîne de traitements de données aux sources différentes. L'utilisabilité de jeux de données ne respectant pas le standard majoritaire est faible. En effet, la mise en correspondance d'un jeu de données au format numérique ancien

[19] Traduction par l'ISO de « *the extent to which a system, product or service can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use* ». consulté en ligne en avril 2019 : https://www.iso.org/obp/ui/#iso:std:iso:9241:-11:ed-2:v1:fr

ou à la structure fortement distincte exige un investissement en temps de traitement pour sa mise en conformité (Brown et *al.*, 2013).

La première étape du partage des données, franchie au début des années 2000, a montré l'importance de la formalisation des modèles de données. Cependant, malgré ces formalisations, l'intégration de différentes sources de données en un même système reste longue et coûteuse, faute de normalisation. La normalisation, au moins conceptuelle, est devenue un passage nécessaire à l'accomplissement de certaines tâches automatiques de raisonnement qui mobilisent un certain nombre de connaissances (Charlet et *al.*, 2004).

La problématique de l'« utilisabilité » d'une base de données est posée de manière plus pressante par l'incitation à l'ouverture des données et par les opportunités techniques. Celles-ci exigent une interopérabilité entre les multiples bases de données déjà constituées dans le monde. En outre, dans le domaine de l'information géographique, cette interopérabilité attendue vise à la constitution de ressources documentaires couvrant le monde, à différentes échelles et sur des thématiques variées. Pour être interopérables, les bases de données doivent partager de manière consensuelle et formalisée des terminologies, des définitions et des concepts qui constituent la part de la réalité décrite en commun.

L'essor des systèmes d'informations géographiques et le développement de méthodes informatiques d'analyses spatiales, croisant des couches numériques d'informations, renforcent le besoin de mutualiser des bases de données. Ce qui, dans les années 1990, était techniquement inconcevable est devenu une problématique du monde de l'écriture numérique. Cette problématique reconnue, notamment au travers du concept d'interopérabilité, s'incarne dans la définition de formalismes d'écriture informatique. Ce formalisme a pour objectif majeur la standardisation des formes d'enregistrement des données, avec pour objectif le partage de données entre les services.

1.1.2. Pour l'interopérabilité des bases de données géographiques

Le premier maillon de la chaîne de production de données géographiques est l'étape conceptuelle par laquelle, dans un cadre théorique, une sélection d'entités géographiques est prévue pour être intégrée dans une base de données afin de répondre

à des besoins expérimentaux ou gestionnaires, par exemple. Afin de garantir la production et l'« utilisabilité » d'une base de données, dans son contexte d'utilisation spécifié, le processus menant à sa production doit être explicité ou formalisé en amont. Cette phase explicative débute par la délimitation de la part de la réalité, qui sera renseignée par la représentation en données géographiques.

Un des cadres théoriques mobilisable pour la production, peut être désigné par le terme d'ontologie (Gruber, 1995 ; Monnin et Félix, 2009 ; Bachimont, 2000). L'ontologie apporte un cadre consensuel (Charlet.et *al.*, 2004 ; Bachimont, 2006) à partir duquel une abstraction de la réalité est développée en tant que modèle conceptuel. Le terme ontologie renvoie à la fois à une méthode appliquée dans les sciences de l'information[20] et, par métonymie, aux résultats ainsi obtenus. Dans le sens qui nous intéresse, ce concept est issu des sciences de l'information et ne doit pas être confondu, bien qu'il en partage la pratique de la classification et l'objectif d'exhaustivité, avec son emploi en philosophie. Le terme ne désigne d'ailleurs pas seulement un concept : en informatique appliquée, il renvoie à une pratique d'inventaire des différentes dimensions de la réalité (Monnin et Félix, 2009) et s'intègre dans un abondant lexique propre aux systèmes d'informations. L'ontologie est, dans ce cas, le résultat consensuel d'une recherche sur les définitions de concepts logiquement associés. Une ontologie est constituée d'un ensemble de définitions et d'axiomes, qui explicitent la signification de chaque élément de la réalité pris en compte ainsi que leurs interrelations. Cette conceptualisation formelle du périmètre décrit dans la base de données est une ontologie. Une fois les concepts assemblés avec cohérence, ceux-ci forment un socle commun, permettant d'intégrer les nouvelles connaissances du domaine, structurées selon des ontologies (Bachimont, 2006).

On peut distinguer, en réponse à notre besoin cartographique, quatre niveaux d'ontologie : l'ontologie générique, mobilisée comme socle commun, qui est un cadre conceptuel et méthodologique au sein duquel chaque concept est formalisé et identifié ; l'ontologie du domaine, qui rassemble les définitions spécialisées d'un domaine ; l'ontologie d'application, qui est un catalogage d'entités d'une portion de la réalité ; l'ontologie formalisée, qui est la représentation d'une conceptualisation du domaine

[20] Inspirée du domaine propre à la philosophie classique et définie par le Centre National de Ressources Textuelles et Lexicales du CNRS (CNRTL) comme la « partie de la philosophie qui a pour objet l'étude des propriétés les plus générales de l'être, telles que l'existence, la possibilité, la durée, le devenir » (ontologie, CNRTL), consulté en ligne en janvier 2019 : https://www.cnrtl.fr/definition/ontologie.

applicatif dans les termes d'un langage informatique opérationnel[21]. La figure 1 représente les propriétés des quatre types d'ontologies cités.

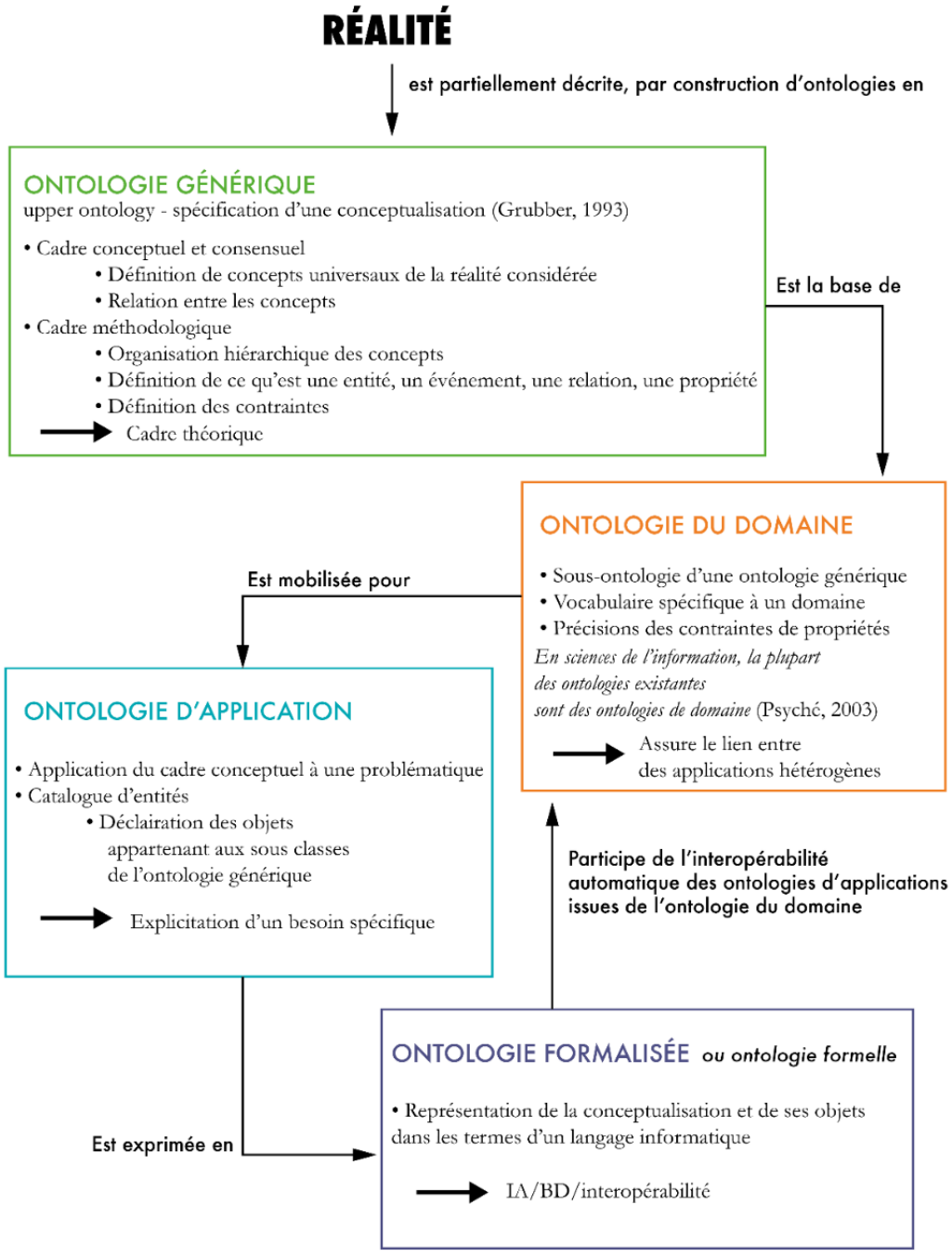

Figure 1 Exemple d'utilisation de la méthode des ontologies pour la construction de bases de données (Hayat, 2019)

[21] Cette distinction en quatre niveaux d'abstraction des pratiques de l'ontologie s'inspire des propositions de typologies des ontologies proposées par Thomas Gruber, Alexandre Monnin et Edith Félix et Jean Charlet, Bruno Bachimont, et Raphaël Troncy.

Dans cette sélection de types d'ontologie, l'ontologie générique est le cadre conceptuel le plus abstrait. Il garantit des bases consensuelles. L'ontologie du domaine et l'ontologie d'application sont concentrées sur les concepts nécessaires à une bonne description d'informations géographiques. Ces deux moyens d'explicitation sont fortement liés et on peut considérer que leurs limites sont poreuses. L'ontologie d'un domaine a deux fonctions principales. En premier lieu, ses concepteurs et ses utilisateurs partagent une même compréhension du domaine grâce à la définition des concepts individualisés. En second lieu, l'ontologie du domaine peut devenir un outil au service de l'organisation d'un ensemble d'informations qui, reliées, constituent un réseau de connaissances (Mechouche et *al.*, 2011). Pour constituer le fondement théorique d'une base de données, il est utile de concevoir l'ontologie d'application, spécifique au besoin, qui dépend d'une ontologie du domaine, la géographie dans notre cas, en s'appuyant sur une ontologie générique consensuelle. Ces trois types d'ontologie sont spécifiés dans un langage indépendant de l'implémentation technique et à un niveau d'abstraction détaché de la structure des données (Gruber, 2009). Enfin, on appelle ontologie formelle l'écriture dans un langage informatique du schéma propre à la gestion de bases de connaissances (Bachimont, 2006). Ce nouveau formalisme a permis l'établissement d'une sémantique exploitable par le biais des machines. Mais sa singularité est d'être simultanément directement compréhensible et facilement partageable entre humains. L'ontologie formelle est la grammaire pour une écriture informatique opérationnelle. Cette simultanéité des partages garantit en principe une complète interopérabilité. L'ontologie est qualifiée de formelle dans la mesure où elle explicite un ensemble de règles destinées à l'écriture numérique des connaissances. Elle est décrite dans la norme ISO 19150[22] relative à l'information géographique et définit les règles de conversion des vues statiques UML et des schémas d'application (ou schémas conceptuels) de l'information géographique en ontologies OWL afin d'utiliser et de prendre en charge l'interopérabilité de l'information géographique dans le Web sémantique. L'ontologie formelle est alors le moyen technique, méthodologique et conceptuel qui permet une interopérabilité entre les systèmes d'information, basée sur la sémantique des données. En effet, le web sémantique, en construction, vise à faciliter la navigation en ligne, parmi toute la diversité des expressions de connaissances, grâce

[22] Norme intitulée Information géographique -- Ontologie -- Partie 2: Règles pour le développement d'ontologies dans le langage d'ontologie Web (OWL). Consulté en ligne en avril 2019 : https://www.iso.org/fr/standard/57466.html?browse=tc

à des liens sémantiques. Ces liens sont conçus comme devant être porteurs de sens aussi bien pour une machine que pour un humain.

Dans les sciences de l'information, cette organisation prend la forme d'une base de connaissances dont l'ontologie est le cadre conceptuel et dont une ontologie formelle en est la description fonctionnelle. Une base de connaissances est une ressource en savoirs qui n'est pas figée. Grâce aux relations créées entre les entités et grâce à la normalisation des descriptions suivant des formats numériques, il est possible de croiser les classifications. Ainsi, par exemple, une ontologie du domaine définirait le sens du concept de *ville*, l'associerait au concept générique *d'habitat,* lui-même associé au concept générique de *peuplement*. L'ontologie du domaine présenterait un ensemble de contraintes et de propriétés attendues pour la description d'un objet *ville* unique comme, par exemple, un nombre d'habitants distinguant les sous-classes du concept *ville*[23]. Toujours à titre d'exemple, dans le cas d'une production d'une base de données de villes françaises, on pourrait poser, dans l'ontologie d'application des contraintes à la variable *nombre d'habitants* par *ville* en utilisant les classes posées par l'INSEE. Ce type de classification désormais fortement formalisé, est à l'œuvre depuis la création des bases de données. Mais il ne l'a été, jusqu'à la définition du langage OWL, que de manière empirique. Il trouve dans la réflexion par ontologie le moyen d'être systématisé. Cette nécessité d'un éclaircissement et d'une explicitation théorique est née des malentendus entre producteurs et utilisateurs des bases de données, portant sur les usages informatisés possibles de ces bases.

[23] Exemple de description tiré de la hiérarchisation faite par les contributeurs à Wikidata.

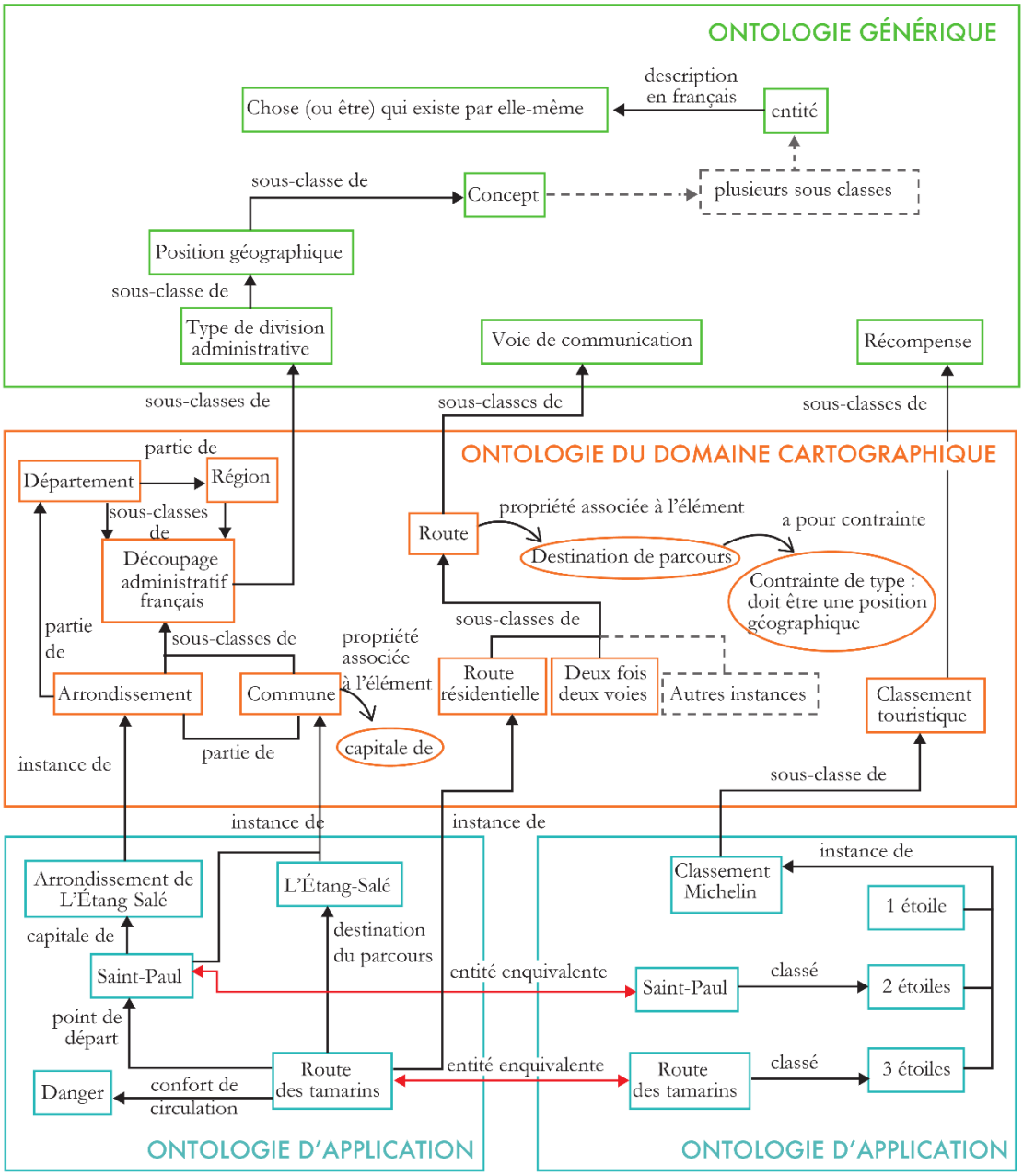

Figure 2 Mobilisation d'ontologies pour la construction d'une base de données cartographique pour la création d'une carte de l'île de La Réunion (Hayat, 2019)

La figure 2 montre que pour deux applications (à gauche, la construction d'itinéraires à la Réunion et à droite la classification de lieux selon leurs qualités pittoresques) exigeant des informations différentes sur le territoire de l'île de la Réunion, il est possible de concevoir des ontologies d'application relevant du même cadre conceptuel. Dans cet exemple, la route des tamarins est une voie de communication dont les points de départ et de destination doivent être déclarés. Les propriétés doivent être des lieux : il s'agit, dans notre exemple, des communes de l'Étang Salé et de Saint Paul. Cette

formalisation mettant en correspondance ontologique deux objets permet ainsi de concevoir un schéma conceptuel général d'une base de données répondant au besoin de création d'une carte complexe, mobilisant différentes ontologies d'application comme la carte touristique de l'île de la Réunion.

Si le Web et l'informatisation encouragent à davantage d'explicitations lors de l'enregistrement de l'information, ces deux innovations ont également mené à l'émergence de la *folksonomie*. Proposée en 2004 par un expert en architecture de l'information (Thomas Vandel Wal). **Le terme *folksonomie* désigne la pratique de classification d'informations (photographies, pages web, signets) par les utilisateurs d'Internet qui marquent des ressources disponibles sur le réseau par des étiquettes (ou *tags*)**. Les termes de classification et d'usagers sont les clés du processus décrit par l'expression de folksonomie puisque cette dernière est constituée de la combinaison des mots « *taxonomy* » (classification) et « *folk* » (usagers). Le rôle d'Internet dans cette pratique est fondamental. C'est en outre, les opportunités d'interactions, propres à l'évolution du web en un web interactif, qui rendent possible la création en ligne par les utilisateurs. Ainsi, les internautes ne sont plus uniquement utilisateurs mais également producteurs. À propos de cette ambivalence, deux mot-valise émergent pour décrire les internautes : les *prosumers* qui sont des professionnels-producteurs-consommateurs et les *produsers* qui sont des professionnels-producteurs-utilisateurs).

Le terme de *prosumers* est proposé, en 1980, par le sociologue Alvin Toffler dans le livre *The Third Wave*. Le concept n'est donc pas issu d'observations de comportements en ligne. Le *prosumers* consiste en la participation des consommateurs au processus de production. Les industries offrent aux consommateurs la possibilité de personnaliser les produits qu'ils souhaitent acheter. Par cette opportunité de personnalisation et d'adaptation des produits aux besoins des consommateurs, les producteurs acquièrent des connaissances sur les attentes des clients. Ceux-ci ont une marge modérée d'expression de leurs besoins ou de leurs créations qui sont contrôlées par les producteurs. Au contraire, l'expression *produsers* (Bruns, 2006), décrit un modèle de production (tourné vers la production de connaissances) selon lequel : « *the production of ideas takes place in a collaborative, participatory environment which breaks down the boundaries between producers and consumers and instead enables all participants to be users as well as producers of information and knowledge, or what [Bruns] have*

come to call produsers »[24] (Bruns, 2006). Ainsi, si le modèle de *prosumers* persiste en ligne, dans l'exploitation des données personnelles par les industriels comme Amazon ou Google (Bruns, 2013), une autre forme de création, non guidée se développe en ligne et est pratiquée par des *produsers* ou ce que nous appelons dans ce travail des contributeurs.

Les contributeurs en ligne produisent des informations dans le but de les rendre utilisables ou de les utiliser par eux-mêmes. Cette production peut se pratiquer par la folksonomie. La folksonomie est d'abord une méthode de description qui repose sur des mots-clés. Par exemple, sur Flickr on attribuera les mots-clés « aventure », « cinéma », « la Butte-aux-Cailles » à des photos. De cette manière il est possible d'organiser les photos que l'on a ainsi étiquetées pour un usage personnel[25]. Cependant, l'usage des mots-clés, plébiscité pour sa simplicité d'utilisation, est également choisi dans certains contextes de production contributifs dont la portée est collective. La création d'informations par la folksonomie présente donc un enjeu fort d'interopérabilité. Alors que les contributeurs sont libres de choisir les mots-clés qu'ils souhaitent, ils visent en même temps à la constitution de bases de données collectives. Ainsi et alors que paradoxalement la folksonomie a été choisie pour construire des projets contributifs, cette méthode de description ne prévoit pas de manière intrinsèque le consensus, pourtant utile à la construction d'un produit collectif. De ce fait, la folksonomie est une pratique largement mobilisée en ligne et qui vient percuter la solution de l'ontologie.

Ces deux processus de description de l'information, folksonomie et ontologie, visent tous deux à soutenir l'exploration de grandes quantités d'informations dans le souci de prendre en compte différents points de vue quant à une information. La folksonomie permet au plus grand nombre de décrire une information telle qu'il la perçoit. Tandis que l'ontologie reste une méthode qui soutient l'expression d'un consensus. Ces deux méthodes ne s'opposent pas et chacune inspire la mise en pratique de l'autre. À ces deux méthodes, il faut en associer une troisième, celle de la modélisation. Celle-ci est postérieure à l'ontologie dans le processus de production et d'exploitation

[24] La production d'idées se déroule dans un contexte collaboratif et participatif qui brise les frontières entre producteurs et consommateurs et permet à tous les participants d'être à la fois des utilisateurs et des producteurs d'informations et de connaissances, ou ce que [Bruns] appelle les *produsers*. Traduction de l'auteure

[25] Les termes mots-clés et étiquettes sont des traductions du nom *tag* et du verbe to *tag* que nous emploierons lorsque la discussion porte sur l'implémentation technique de projets contributifs qui choisissent ce terme même entre francophones.

d'informations. Quant à la folksonomie, une réorganisation hiérarchisée des objets étiquetés (par ontologie, par exemple) est nécessaire à sa transformation en un modèle (fig. 3, p.41).

Suite à des analyses, il est possible d'organiser, a posteriori, une folksonomie en une ontologie, si celle-ci s'avère plus pertinente. La transformation d'une folksonomie en une ontologie repose alors sur l'interprétation des expressions choisies par les *produsers*. En effet, alors que les folksonomies sont « des organisations de ressources issues des usages et [que] leur obtention est non supervisée », elles « ouvrent de nombreuses possibilités d'exploitation, pourvu qu'un système de partage efficace des connaissances et des ressources en tire profit » (Limpens et *al.*, 2008). Pour cela, il s'agit de clarifier les *tags*. **En effet, un *tag* peut désigner plusieurs concepts et un même concept peut être désigné par différents *tags***. En plus de ces redondances et ambiguïtés possibles, les erreurs typographiques constituent un frein à la reconnaissance automatique d'informations non indexées par des moyens plus formels. Différentes méthodes visent l'alignement d'une folksonomie sur une ontologie ou la construction d'une ontologie à partir d'une folksonomie. Le niveau de généralisation de l'ontologie dépend des ressources catégorisées par la folksonomie. Cependant, dans un souci d'interopérabilité, plus l'ontologie sera générique, plus les ressources taguées seront mobilisables par des ontologies d'application ou formelles diverses dont les expressions peuvent différer mais qui se rapportent aux mêmes objets d'une ontologie générique commune. Les méthodes de construction d'ontologies peuvent reposer sur un travail de catégorisation essentiellement manuel qui mobilise les techniques du *crowdsourcing*. Par une série de questions, correspondant à un concept de l'ontologie, posées aux contributeurs via une plateforme en ligne ou une application mobile, on dissimule la complexité de l'ontologie afin de permettre à l'utilisateur d'associer un objet à une catégorie (Limpens et *al.*, 2008). Une autre méthode consiste à partir d'une ontologie et, par reconnaissance de chaîne de caractères, à construire des correspondances entre les *tags* de la folksonomie et les concepts de l'ontologie. La reconnaissance de chaînes de caractères présente cependant des limites. Ainsi, par exemple, selon la finesse de l'ontologie, le *tag* « New York City » sera bien identifié comme catégorisant un objet en rapport avec la ville de New York ; mais cela ne sera pas forcément le cas pour le *tag* « Big Apple ». Cette méthode peut alors être couplée

avec une reprise manuelle, en soumettant les identifications automatiques à un utilisateur qui prend une décision (Hombiat, 2017).

Enfin, des méthodes complètement automatiques peuvent être mobilisées pour aligner une folksonomie sur une ontologie. Dans la voie automatique, on regroupe des *tags* par leur proximité orthographique ou par leur cooccurrence. Les regroupements peuvent être confrontés à des ontologies structurées en base de connaissances. C'est l'opportunité offerte par l'interopérabilité du Web sémantique (Hombiat, 2017). Dans la figure 3 de la page suivante, nous présentons des relations qui peuvent être construites entre des ressources en ligne et une ontologie. Nous considérons trois ressources. La première est l'objet décrivant la région Bretagne dans la base de données du projet OpenStreetMap. La deuxième ressource provient également de la base de données OpenStreetMap et décrit, en une relation plus complexe, la suite de sentiers de randonnées reliant le Mont-Saint-Michel au parvis de Notre-Dame à Paris. La troisième ressource est une photographie du Mont-Saint-Michel et de ses alentours. Pour l'exemple, nous n'avons retenu que quelques *tags* mais ces objets ont été catégorisés par plusieurs dizaines de mots-clés. Le *tag* « `admin_level=4` » signifie que l'objet 102740 de la base de données OSM est une région. Il est donc possible de rattacher cet objet au concept de région de l'ontologie du domaine. La combinaison des *tags* « `admin_level=4` » et « `name=Bretagne` » pourrait en outre mener à créer l'instance correspondant à la région Bretagne. Enfin, les *tags* « `wikidata=Q12130` » et « `wikipedia=fr: Région Bretagne` » confirment cette association. En effet, l'encyclopédie et la base de connaissance sont deux moyens du web de désambiguïsation des informations[26]. La photo enregistrée sur la plateforme de partage Flickr porte davantage d'ambiguïté : il semble que la scène se déroule à la fois en Bretagne et en Normandie ! Les procédures automatiques ne peuvent trancher que par un recoupement d'informations. On acquiert cependant l'information que le lieu est classé au patrimoine mondial de l'UNESCO. Compte tenu de notre ontologie du domaine, on peut l'associer à un classement touristique. Quant à la seconde relation issue d'OSM, il s'agit d'une succession d'objets représentant des voies de communication pédestres et routières, qui en étant contiguës forment un itinéraire de

[26] Dans le cadre de la pratique d'affectation libre d'étiquettes (*tags*) par les internautes, un *tag* peut désigner plusieurs concepts et un même concept peut être désigné par différents *tags*. Selon l'usage qui est fait de la classification des ressources, il s'agit alors d'identifier des correspondances entre les étiquettes que l'on souhaite assimiler et ainsi confirmer la distinction entre les étiquettes porteuses de sens que l'on souhaite garder distinctes.

randonnée comme l'indique le *tag* « `route=hiking` ». Nous assimilons cet objet à une instance d'une sous-classe de la classe route.

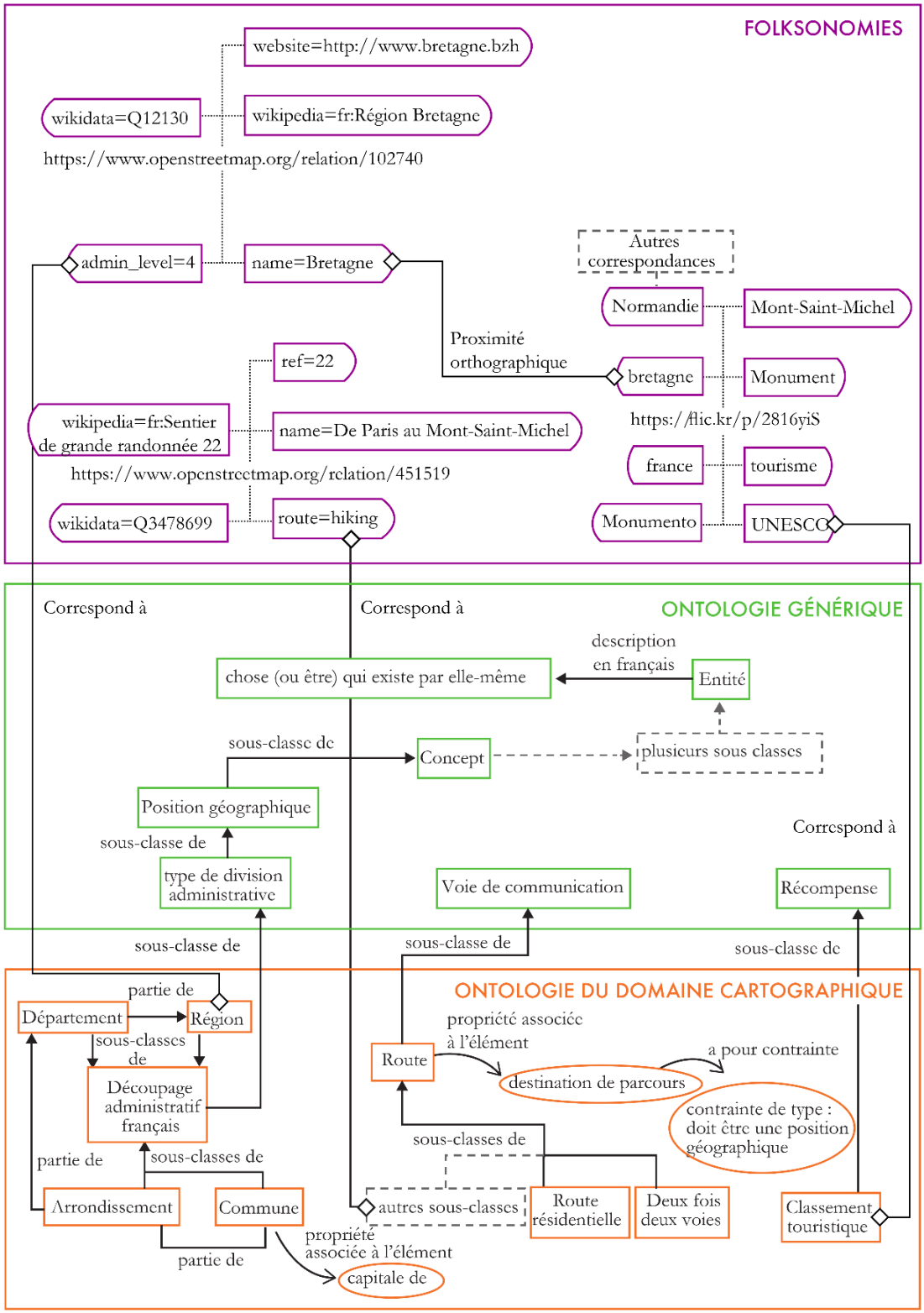

Figure 3 Extraits de tags associés à trois ressources numériques, deux accessible via OpenStreetMap et une via Flickr (Hayat, 2019)

Nous centrons cet exemple sur la réalisation d'une ontologie du domaine cartographique. L'objectif est de construire un cadre théorique, clair et partageable, pour la réalisation d'une base de données cartographique. Dans le contexte de la production d'informations géographiques, tout comme dans celui de la production cartographique, la modélisation est un procédé méthodologique dont le fondement théorique doit être consensuel et explicite. En effet, les utilisateurs doivent pouvoir compter sur un cadre théorique de référence rassurant, intuitif et interopérable. Dans le cadre d'une production cartographique fondée sur une base de données géographique, il est nécessaire, pour les utilisateurs, que soit explicité, en plus du contexte de production technique, le fondement théorique de la production. C'est la théorie, au fondement de la construction de la base de données, qui a permis de classer le réel en catégories et d'en sélectionner les éléments nécessaires aux usages de la base de données. La figure 4 (page suivante) présente, à la suite, soit de la figure 2, soit de la figure 3, la mise en commun de deux ontologies d'application (la première est celle d'un réseau de villes connectées par un réseau routier et la seconde celle de la classification de lieux touristiques) pour la modélisation d'une base de données cartographique.

La modélisation cartographique consiste en un ensemble de choix. Dans le cas présent, on a conservé, suite à une réflexion éditoriale prenant en compte les informations nécessaires à l'utilisateur de la carte pour qu'il se repère et l'échelle de représentation, le découpage administratif jusqu'au maillage communal, les toponymes associés à ce maillage et l'ensemble du réseau routier. La géométrie de ces objets est également modélisée ; ainsi les territoires administratifs devront être dessinés sous forme de polygones. Les toponymes sont à considérer comme des nœuds et les routes des lignes. Enfin, des contraintes logiques sont définies afin de garantir leur utilisation informatisée (les objets doivent être interprétés dans une base de données) mais également la qualité graphique finale. De ce fait, les polygones doivent partager des limites communes et les tronçons de route sont connexes.

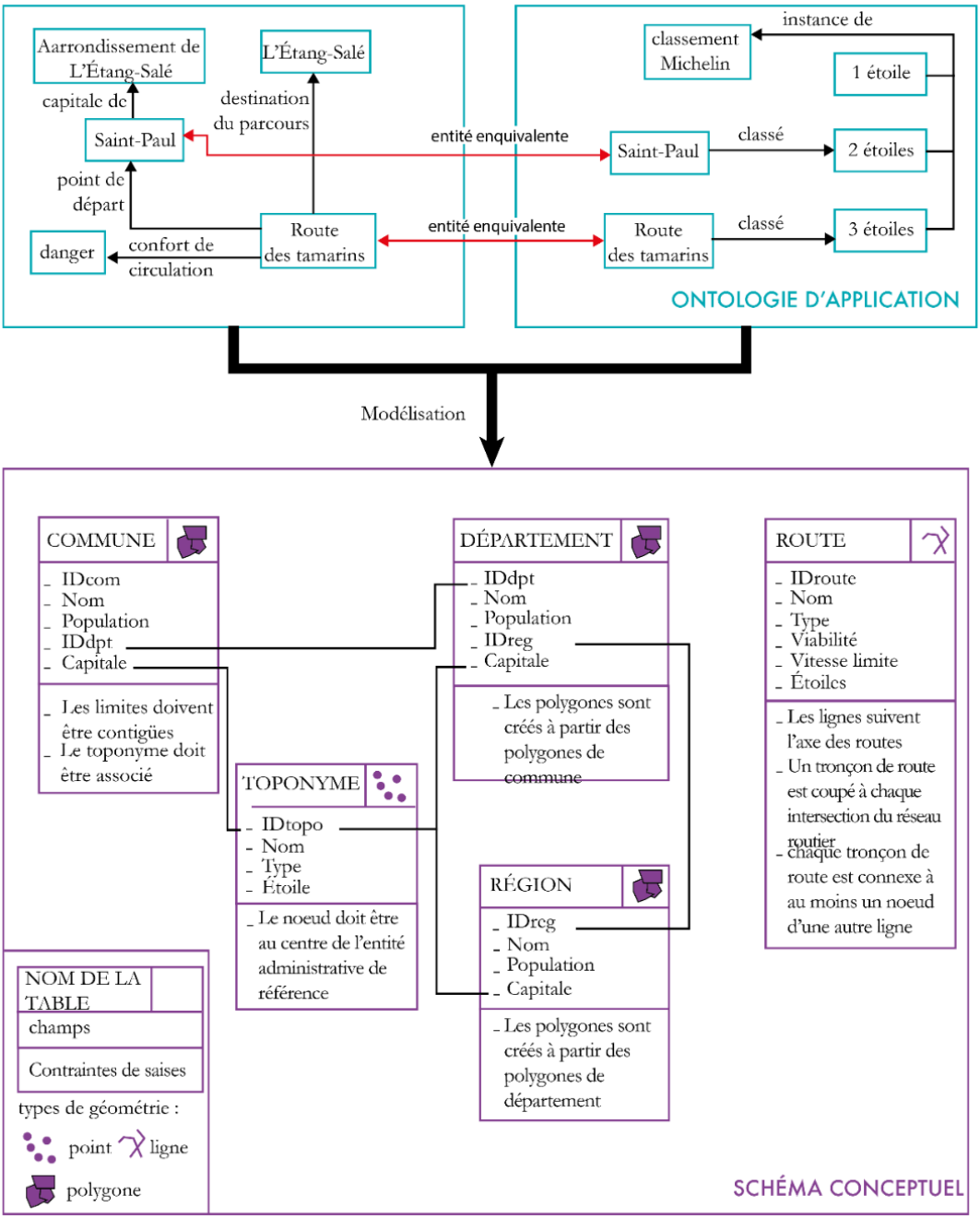

Figure 4 Exemple d'extrait d'un schéma conceptuel issu de la méthode de construction par ontologie (Hayat, 2019)

Les méthodes d'explicitation du processus de création et d'organisation d'informations présentent pour nous un caractère fondamental. Les nouveaux besoins de l'*open data*, du contributif et du web sémantique, au service d'analyses de données automatiques ou d'interopérabilité, exigent désormais du producteur (traditionnel ou *produser*) qu'il formalise son cadre théorique selon des règles accessibles aux êtres humains et aux

machines. L’information géographique devient volontaire contributive et libre mais ce n’est pas sans contraintes.

1.2. L'information géographique volontaire, contributive et libre

À partir de 2007, l'expression d'Information Géographique Volontaire (*Volunteered Geographic Information* –VGI–, Goodchild, 2007) se diffuse dans la littérature scientifique. L'adjectif *volontaire* est proposé pour qualifier le projet OpenStreetMap. Plus largement, cette expression a été reprise par la communauté scientifique pour désigner tout projet de création de bases de données géographiques, dont les producteurs seraient des internautes anonymes. Si le qualificatif de « volontaire » souligne le rôle actif des internautes – ce qui est alors une nouveauté – il ne révèle ni le *caractère contributif,* ni le *caractère libre* de la diffusion ou de la conception des données.

Désormais, l'expression de VGI est complétée par celle de « contenus générés par les utilisateurs » (CGU) et celle de « citoyens capteurs ». Ces deux dernières expressions désignent des situations bien plus diverses que la VGI car elles n'évoquent pas le caractère volontaire de la contribution. En effet, si le CGU correspond surtout aux tweets et aux commentaires sur TripAdvisor volontairement publiés, des informations involontaires s'y ajoutent comme la localisation de la personne au moment où elle participe à ces réseaux en ligne. Certes, cette dernière fonctionnalité doit être choisie sur Twitter, mais d'autres plateformes construisent des bases de données uniquement issues de ces enregistrements de données partiellement involontaires. C'est pourquoi, en complément du terme de VGI, la littérature scientifique propose désormais le *Un-Volunteered Geographic Information*.

Les trois expressions que sont *l'information volontaire*, *le contenu généré par des utilisateurs* et le *citoyen capteur* appartiennent au domaine de la recherche mais elles ne sont pas communément utilisées par les groupes d'individus auxquels elles réfèrent. En effet, le citoyen capteur, producteur d'informations volontaires réalise son activité le plus souvent en groupe. Et quand il parle de ses productions, il s'intègre à une « communauté ».

Les projets contributifs comme Wikimedia Foundation et OpenStreetMap se décrivent, via leurs sites Web et leur *wiki*, comme « *free* ». Ce terme se comprend en un sens restreint : il renvoie seulement à la notion de liberté et non à l'idée de gratuité. D'ailleurs, dans un texte rédigé en anglais, le terme de libre est occasionnellement cité

en français pour sa capacité à référer seulement à l'idée de liberté. À la suite du mot liberté, d'autres noms sont apparus pour préciser les principes de ce qui est devenu le courant du Libre. En effet, les discours de communautés – comme Wikimedia Foundation, OpenStreetMap et d'autres – façonnés au cours de ces vingt dernières années, diffusent, outre le terme de liberté, les mots « ouvert », « partage » et « connaissance »[27]. Les deux termes libre et ouvert sont polysémiques. Ils se rapportent à la liberté d'accès à la connaissance et peuvent être assimilés au mot « partage ». La liberté en question équivaut à une liberté de participer à la création de la connaissance. C'est d'ailleurs ce sens qu'évoque le mot « volontaire » compris dans l'acronyme VGI choisi par Michael Goodchild (2007) pour désigner ces créations. Il est légitime pour un chercheur de pointer l'originalité du mode de production libre, chacun pouvant modifier un document auquel il a le libre accès.

1.2.1. Le Libre et les communs pour l'information géographique

« Une œuvre libre ou contenu libre, est une œuvre de l'esprit dont la diffusion et la modification sont libres. Ces œuvres sont notamment des images, des textes, de la musique et des logiciels dont chacun peut distribuer autant de copies qu'il le souhaite, et aussi les modifier pour les améliorer. La notion d'œuvre libre assure donc la liberté d'expression seulement dans la mesure où elle permet la liberté de diffusion et de modification des informations contenues dans les œuvres et le cas échéant, la commercialisation. » (Œuvre libre. *Wikipédia, L'encyclopédie libre*, 2018). Une œuvre libre est produite selon les principes du Libre. Ce terme fait référence au courant de pensée dont l'essentiel de la formalisation repose sur les premières licences de logiciels, dites libres. Le logiciel est libre car une licence lui assure une protection originale. Cette protection juridique vise à garantir l'annulation des droits exclusifs (copyright ou droit d'auteur) attribués à l'auteur du code (Mangolte, 2015). Cette invention juridique est une alternative aux lois américaines des années 1980, appliquant de manière absolue et systématique le copyright sur une œuvre nouvellement créée. Le concept à la source de ce don de liberté est le suivant : un groupe peut entretenir une ressource, la protéger et créer de nouvelles ressources en exploitant la ressource originale ou les suivantes. Cette

[27] Le terme connaissance est à rapprocher du terme information. Le choix de ce mot s'explique par la diversité des contenus produits dans le cadre des projets libres, qui ne correspondent pas uniquement à la notion d'information. Une portée universaliste est également visée dans ces projets et le terme de connaissance en est l'expression.

dynamique particulière de création et d'accumulation des savoirs mène à des innovations (Mangolte, 2015) et ainsi à la création de nouveaux savoirs.

Des organisations comme la Free Software Foundation (FSF)[28], la Wikimedia Foundation[29] ou l'OpenStreetMap Foundation[30] sont des incarnations des projets libres et communautaires dont elles portent le nom. Le développement de telles organisations tient au succès de leurs projets, en termes de nombre de contributeurs et de réutilisations des logiciels ou de contenus créés dans le cadre des projets. Ces organisations ont influencé globalement la pratique de la production de connaissances libres via Internet en proposant des règles de partage des contenus, des codes informatiques et des bases de données. Elles soutiennent de nombreux projets contributifs et diffusent des argumentaires en faveur du Libre. Ainsi, la Free Software Foundation se décrit-elle en première page de son site Web comme « *a nonprofit [organization] with a worldwide mission to promote computer user freedom* »[31]. Dans le cas de la Wikimédia Foundation, ce n'est pas le logiciel qui est mis en évidence en première page mais la ressource qu'est la connaissance. Le Libre (*free*) est mentionné comme suit : « *Imagine a world in which every single human being can freely share in the sum of all knowledge* ». De manière très similaire, l'OpenStreetMap Foundation se décrit comme « *an initiative to create and provide free geographic data, such as street maps, to anyone* ». Les notions de libre accès et de création par tous sont communes à ces trois grands acteurs du courant Libre.

Si la notion de communauté est évoquée à la suite de chacun de ces messages de présentation, l'organisation communautaire reste très lâche dans la pratique. Nombreux sont les contributeurs anonymes. D'ailleurs, les modes de production mis en place n'exigent pas des liens forts entre les contributeurs. Seul le respect de règles propres aux projets est attendu de tous les contributeurs. Ces règles sont posées par les volontaires au moment de leur rédaction et elles sont débattues publiquement. Ainsi,

[28] La Free Software Foundation est une organisation à but non lucratif, fondée dans le but de promouvoir la conception de logiciels selon les principes du Libre. Fondée en 1985 par Richard Stallman, l'institution est une figure du mouvement Libre en tant que principal et premier diffuseur de ses concepts.

[29] Institution à but non lucratif, fondée en 2003, Wikimédia Commons a été fondée afin d'administrer les projets comme Wikipédia, Wikitionnaire ou Wikidata.

[30] La Fondation OpenStreetMap est la personne morale permettant au projet OpenStreetMap de collecter des fonds et de présenter des représentants auprès d'institutions et d'acteurs privés. Fondée en 2006, sa création a eu pour principale motivation de ne pas conserver uniquement le nom de Steve Coast comme responsable du nom de domaine et des serveurs de stockage.

[31] La Free software Foundation se décrit comme une organisation à but non lucratif ayant pour mission de promouvoir la liberté des utilisateurs de matériels informatiques. Traduction de l'auteure.

les débats liminaires réunissent-ils un nombre réduit d'individus et les règles dont ils ont convenu sont ensuite tacitement admises par les non débatteurs. Les groupes producteurs d'informations volontaires que nous venons d'évoquer se conçoivent comme des acteurs du Libre. Cependant, d'après les théories de l'économiste américaine Elinor Ostrom – Prix Nobel 2009 – relatives à l'exploitation d'une ressource en groupe, nous pouvons aussi décrire leur activité comme un fonctionnement en « commun » (*commons*).

En 1973, Elinor Ostrom et son époux, Vincent, ont fondé L'Atelier de théorie et d'analyse des politiques (*Workshop in Political Theory and Policy Analysis*) à l'université d'Indiana à Bloomington (États-Unis). Ils ont mené, en collaboration avec leurs étudiants et avec des chercheurs du monde entier, des expérimentations qui tendent à démontrer la viabilité des organisations en commun et à analyser le potentiel de participation des citoyens à plusieurs niveaux de la société, du très local au très global. Leurs conclusions, issues de quarante ans de travaux de recherches pluridisciplinaires, contredisent les thèses des défenseurs du concept de l'*homo-economicus*. Elles invalident deux conceptions pessimistes de la citoyenneté. La première, défendue par Garrett Hardin dans *The Tragedy of the Commons* (1968), considère l'homme comme un être avide. Selon cette conception, chacun cherche à tirer profit des ressources du bien commun tout en s'exemptant de les enrichir, aux dépens de leur pérennité. La seconde conception est soutenue par Mancur Olson dans *The logic of collective action* (1965). D'après lui, dans les organisations de grande dimension, les individus ont tendance à profiter de l'action collective de création de la ressource, depuis l'extérieur, sans y apporter la moindre contribution. Mancur Olson est le premier à avoir nommé ce phénomène « *free-riding* » ou « passager-clandestin ». Ces deux auteurs concluent leurs raisonnements en expliquant qu'en situation de partage de biens communs, chacun tend à sa propre ruine et à celle de tous, de sorte que tout bien devrait être soit privatisé, soit nationalisé. Dans les deux cas, des inégalités paraissent inévitables. Mais d'après Garrett Hardin, seules la privatisation ou la nationalisation permettent aux hommes de se préserver de « la tragédie ». À ces conclusions pessimistes, Elinor Ostrom substitue une méthode expérimentale et systémique.

Les « communs » ont pour visée la gouvernance[32] d'un ensemble de ressources dans le but de permettre un accès partagé à ces ressources (Coriat, 2015, p.13). Le nom générique de « communs » a été le sujet de recherche de l'économiste Elinor Ostom depuis 1990. Elinor Ostrom et ses collaborateurs ont repris « un vieux mot et une tradition ancienne » (Coriat, 2015, p.10) et les ont remis sur le devant de la scène scientifique, montrant que ce mode d'organisation n'était pas abandonné. Après dix-sept ans d'études sur les communs fonciers comme les pâturages, les pêcheries ou l'accès à l'eau et leur gouvernance, Elinor Ostrom s'est tournée en 2007 avec Charlotte Hess vers ce qu'elles ont nommé les communs de la connaissance, regroupant les communs informationnels, virtuels ou numériques. À la suite des travaux d'Ostrom et Hess, depuis les années 2000, dans les domaines du droit, de la sociologie, de l'économie, des sciences de l'information, la recherche s'est mise à comparer les communs tangibles avec les communs informationnels afin de proposer une grille d'analyse de ces nouvelles formes de commun (Coriat, 2015 ; Laval, 2016 ; Macbeth et Pitt, 2015 ; Hess et Ostrom, 2007 ; Frischmann et *al*., 2014).

La logique des communs n'exclut ni le marché ni l'État. Mais elle peut correspondre plus efficacement à certains besoins. Une fois le modèle des biens communs repéré et adapté à la gestion de la ressource, une gouvernance est à établir afin de garantir son caractère commun. Dans *Governing Commons*, Elinor Ostrom met en évidence huit critères qui correspondraient à ces mécanismes constitutifs d'une organisation en commun (Ostrom 1990, 90–102). Ces principes sont les résultats d'observations des « modes d'organisations des biens communs à une échelle petite et moyenne dont certains avaient réussi et d'autres avaient échoué ». Elinor Ostrom a tenté de poser les principes permettant de comprendre les échecs et les succès de divers communs. Ces principes ne sont pas à comprendre comme les fondements d'un nouveau dogme ou comme un programme figé de construction d'une nouvelle société, mais comme des outils conceptuels à disposition des communautés pour qu'elles établissent elles-mêmes leurs règles de gouvernance. « Lorsque vous envisagez de mettre au point un nouveau système, déclare Elinor Ostrom, il est utile de réfléchir aux contours d'une ressource, de façon à ce que les gens puissent savoir exactement quelles sont celles dont ils disposent, où ils peuvent chasser, prélever du poisson, abattre des arbres » (Ostrom,

[32] Au sens d'Elinor Ostrom c'est-à-dire : « un système de règles et normes, définissant les droits (et obligations) attribués aux différents membres d'une communauté » (Oliver Weinstein dans Orsi et *al.*, 2017).

2011). Il est généralement facile de retrouver une application de la logique de la mise en commun, lorsque celle-ci opère à un niveau local, comme cela fut longtemps réalisé, par exemple, autour d'un moulin ou de jardins partagés. Mais qu'en est-il d'un « commun » comme le climat ? La question ne sera pas traitée dans ce travail mais son ampleur globale équivaut à celle des biens communs du numérique.

Les huit principes favorisant le succès d'un commun, tels que nous les avons traduits d'après l'ouvrage *Governing Commons* (Ostrom 1990, 90–102)[33], sont les suivants :

- Des frontières clairement définies pour chacun des membres quant à leurs droits sur la ressource,
- L'adéquation entre les règles d'accès et d'utilisation de la ressource et les circonstances propres au commun,
- Des choix collectifs permettant la participation de la majorité des usagers au processus de prise de décision,
- Une évaluation efficace et continue par des agents qui font partie du groupe des usagers ou qui sont responsables devant eux,
- Des sanctions graduées en fonction de la gravité de l'attaque envers la ressource et des connaissances de l'individu sanctionné quant aux règles du commun,
- Des mécanismes de résolution des conflits peu coûteux et facile d'accès,
- La reconnaissance de ces modes d'organisation par des structures supérieures, par exemple les instances gouvernementales,
- Pour les biens communs de grande dimension : une organisation en poupées gigognes, avec des unités imbriquées jusqu'au niveau local.

Ces huit principes d'agencement en communs sont des propositions de mode de fonctionnement servant à se prémunir des vulnérabilités inhérentes à leur statut. La chercheuse a insisté sur le caractère particulier de chaque commun. Les communs numériques le sont d'autant plus qu'ils sont des biens non-rivaux et non-exclusifs à la différence des autres communs matériels, qui sont des biens rivaux. En économie, la rivalité signifie que l'utilisation d'une ressource par un individu en prive les autres. Les communs numériques restent, comme tous les communs, à la charge d'une communauté responsable de la gouvernance.

Bien qu'elle se soit intéressée à Internet à la fin de sa carrière, Elinor Ostrom n'a pas étudié l'organisation des communautés du Libre à proprement parler. Mais son concept

[33] Proposition de traduction par l'auteure de la thèse.

de « bien commun de la connaissance », issu de ses résultats relatifs aux communs naturels est également pertinent quant aux communautés de contributeurs du Libre. Une communauté a la charge de protéger et de faire croître une ressource qui échappe au droit de propriété exclusif. D'ailleurs, dans la perspective d'Elinor Ostrom, l'expression « communs » suppose une interaction sociale permettant le partage et la préservation de la ressource (Ostrom, 1990). Les systèmes des communs ne supposent pas une absence de gouvernance. « Le groupe de personnes » ou la communauté partagent la ressource selon des règles. Le groupe a la charge de contrôler le respect de ces règles et la résolution d'éventuels conflits. En effet, la ressource est vulnérable et menacée de destruction ou d'appropriation. Or, bien qu'elle soit pertinente, l'expression « biens communs » est quasi absente des argumentaires des acteurs du Libre. Pourquoi ? Il est probable que la neutralité des mots « free » et « libre », considérés par ailleurs comme fédérateurs, ait été préférée à la tonalité politique de l'expression « biens communs ». Pourtant, la notion de « biens communs » présente l'intérêt de désigner à la fois une organisation sociale et un processus de création et de conservation.

Internet a apporté un nouveau contexte au système relativement oublié des communs : les communs numériques et notamment les communs de la connaissance. Dans *Understanding Knowledge as a Commons* (2007) dirigé par Charlotte Hess et Elinor Ostrom, les deux autrices présentent une nouvelle approche de la connaissance[34], considérée comme une ressource partageable et partagée dans un « écosystème complexe » qu'est un commun par un groupe de personne « sujet à des dilemmes sociaux ». Elles décrivent l'apparition d'écosystèmes complexes, les communs de la connaissance, comme survenue « avec une soudaineté frappante » entre des utilisateurs d'Internet. Les communs de la connaissance se sont construits grâce aux opportunités de stockage et de partage au sein de ce réseau hyper connecté. Les chercheuses remarquent également que le partage des connaissances via Internet n'a pas été immédiatement accompagné de systèmes de protections et de gouvernance. Ce n'est qu'après une prise de conscience de la mise en danger de la conservation de ces connaissances[35] et de leur accessibilité constante, suite à des conflits, que les communs

[34] « *Knowledge in this book refers to all intelligible ideas, information, and data in whatever form in which it is expressed or obtained* » (Hess et Ostrom, 2007, p.7)

[35] « *People started to notice behaviors and conditions on the web—congestion, free riding, conflict, overuse, and « pollution»—that had long been identified with other types of commons. They began to notice that this new conduit of distributing information was neither a private nor strictly a public resource* » (Hess et Ostrom, 2007, p.4)

au sens de systèmes se sont constitués[36]. Comme Charlotte Hess et Elinor Ostrom le signalent, différentes expressions sont mobilisées pour décrire des systèmes proches : communs de l'information, communs virtuels, communs intellectuels ou communs numériques. Chacune de ces expressions désigne une nouvelle pratique du commun et le concept de commun, au sens restreint où l'emploie Ostrom, n'a pas été adopté par les groupes créateurs de biens communs de la connaissance. Un organe de la Wikimedia Foundation a, certes dans son nom, le mot *commons* : Wikimedia Commons mais, paradoxalement, l'idée du commun n'est pas développée dans sa page de présentation. Le commun y est exclusivement associé à la libre circulation, au libre accès et à la libre édition. L'organisation qui mobilise le plus explicitement le concept de commun, sans toutefois le définir, est l'association « Creative Commons » qui se présente ainsi : « *Help us build a vibrant, collaborative global commons* »[37]. Si ces groupes ne mobilisent pas la terminologie d'Ostrom, c'est en partie par crainte de s'exclure de productions industrielles qui n'adhèrent pas au mouvement social du Libre.

Les pratiques associées de diffusion des connaissances et de repartage des progrès constituent le Libre. Un élément fondamental de ces pratiques est le libre accès à une ressource. Le terme se référant à ce libre accès est celui d'« ouvert » ou « *open* ». Est dite « ouverte » une ressource de nature numérique et structurée selon des formats d'écriture numériques ouverts[38]. L'accessibilité doit être rapide et simple pour tout utilisateur, via un lien partagé sur un site Internet ou sur un portail. Ce lien doit être maintenu afin de rester accessible tant que la donnée est complète, non agrégée et à jour. En outre, le terme « ouvert » exclut la possibilité, pour le diffuseur, de restreindre l'exploitation des données aux seuls abonnés de son service. Le terme « ouvert » découle de la logique du Libre. Une publication ouverte suppose le libre accès mais pas nécessairement la libre édition. Pourtant, une partie des publications en ligne, étiquetées ouvertes, ont un objectif de réutilisation. Ainsi, dans le *Vocabulaire de l'informatique et du droit* par Legifrance (2014) les données ouvertes sont-elles des « données qu'un organisme met à la disposition de tous sous forme de fichiers numériques afin de permettre leur réutilisation »[39]. On retrouve cet usage dans la définition de référence de

[36] Nous développons le processus de construction de protections et de droit d'usage dans la section 1.3 de ce chapitre.
[37] Aidez-nous à créer un commun mondial palpitant et collaboratif. Traduction de l'auteure.
[38] Sur ce qui peut paraître comme une tautologie, voir section 1.3.
[39] JORF n°0103 du 3 mai 2014 page 7639 , consulté en ligne en avril 2019 : https://www.legifrance.gouv.fr/affichTexte.do?cidTexte=JORFTEXT000028890784&categorieLien=id

l'*Open Definiton*[40], diffusée par l'*Open Knowledge Network*. On peut également citer, comme autre consigne de bonne pratique d'ouverture des données, le cas de Tim Berners-Lee. En 2010, l'inventeur du World Wide Web (WWW), décrit une échelle de qualité d'ouverture des données, connue sous le nom des « *5-star Open Data* ». Selon son échelle, un jeu de données ouvert acquiert une étoile pour chacun des aspects suivants : les données doivent être brutes, disponibles selon des formats numériques ouverts (.CSV, .XLM), sous licences ouvertes, accessibles via une page Web et enfin, la structure du jeu de données doit permettre de le lier à d'autres bases de données. Le terme ouvert répandu largement dans le langage courant comme dans le discours des institutions publiques est complémentaire du terme de libre. Le terme *open source* proposé et diffusé à partir de 1998 par l'association *Open Source Initiative* (OSI) vise à se distinguer du courant, considéré comme philosophique et politique du « logiciel libre »[41]. Rejointe par des figures du développement en libre accès et en libre édition, l'OSI est aujourd'hui une référence validant des logiciels dits *open source*, dont une visée est de s'intégrer au système économique des logiciels. En outre, cette scission avec le mouvement du Libre est également actée par la Free Software Foundation. On peut lire dans la page « *Why Open Source misses the point of Free Software* » (Stallman) les phrases suivantes : « les deux expressions décrivent à peu près la même catégorie de logiciels mais elles représentent des points de vue basés sur des valeurs fondamentalement différentes. L'*open source* est une méthodologie de développement ; le logiciel libre est un mouvement de société »[42]. Cette dernière citation montre bien les éventuelles tensions entre projets du Libre et projets *open source*. Néanmoins, le Libre a inspiré un esprit collaboratif qui s'est répandu sur Internet. Olivier le Deuff considère que le Libre a initié une « ouverture d'esprit » sur Internet qui caractérise le comportement actif de certains internautes qui participent à l'indexation de documents en ligne par la création de folksonomies (Le Deuff, 2006). Il rapproche la pratique

[40] « *Open means anyone can freely access, use, modify, and share for any purpose (subject, at most, to requirements that preserve provenance and openness*) » : https://opendefinition.org/. Nous traduisons en : Ouvert signifie que n'importe qui peut librement accéder, utiliser, modifier et partager quelle que soit l'usage (sous réserve, d'exigences qui précèdent l'ouverture).

[41] Récit selon l'Open Source Initiative de l'invention du terme lors d'une conférence en Californie en février 1998 : « *The conferees also believed that it would be useful to have a single label that identified this approach and distinguished it from the philosophically- and politically-focused label " free software*" ». Consulté en ligne : https://opensource.org/history

[42] Traduit de l'anglais « *The two terms describe almost the same category of software, but they stand for views based on fundamentally different values. Open source is a development methodology; free software is a social movement* ». Disponible en français à la même adresse : https://www.gnu.org/philosophy/open-source-misses-the-point.en.html

d'indexation libre de la notion de partage constitutive de la pensée du Libre. Nous sommes en accord avec ce qu'exprime Le Deuff en retrouvant les points communs d'une participation volontaire pour l'amélioration d'un logiciel ou du partage des documents dont le fonctionnement s'appuie sur l'incrémentation des connaissances de tous les participants ou *contributeurs*.

1.2.2. Volontaire et contributif

En 2007, Michael Goodchild a proposé l'expression « information géographique volontaire »[43] (VGI) pour désigner l'engagement volontaire d'un grand nombre de particuliers dans la création d'informations géographiques. Goodchild citait comme exemples « convaincants » les cadres de production de VGI que sont Wikimapia[44] et Flickr[45]. Le premier est le projet de création d'un index géographique qui suit le modèle de production de Wikipédia. Le second est une plateforme de partage de photographies dont l'interface permettait aux particuliers de coupler leur publication à des coordonnées géographiques. Le dernier exemple que citait Michael Goodchild était celui d'OpenStreetMap, qu'il jugeait le plus sophistiqué des trois. L'article qui est devenu une référence en la matière identifie la spécificité du VGI, qui tient à son contexte de production. En effet, M. Goodchild considère que ce sont les avancées technologiques, associées à Internet, qui ont permis la conception d'un tel cadre de production. Au début des années 2000, il était devenu possible de construire des sites Web conçus pour recevoir et diffuser des contenus générés par les utilisateurs (UGC). Dans le domaine de la conception de coproductions entre internautes, les *wiki*s en sont les exemples les plus avancés. On observe que la modération y est *a priori* réduite et qu'elle est le plus souvent exercée *a posteriori*. Cette modération *a posteriori* est d'ailleurs effectuée par les utilisateurs, eux-mêmes producteurs. Une très faible minorité des utilisateurs possède un pouvoir de modération supérieur à celui de la grande majorité. Ce mode de création itératif est rendu possible par les moyens interactifs du Web.

[43] *Volunteered geographic information.*
[44] https://wikimapia.org
[45] https://www.flickr.com/

Les termes utilisés pour désigner les productions volontaires sont nombreux car elles sont apparues dans des domaines différents comme la géographie, les sciences de l'information ou la sociologie. En 2016, dix-sept auteurs (See et *al*., 2016) ont réalisé un recensement de ces terminologies. Les auteurs distinguent vingt-cinq termes apparus ces trente dernières années dans la littérature scientifique, destinés à désigner l'information venue des particuliers. See et *al.* proposent une analyse statistique d'un premier corpus composé de près de 26 000 articles, publiés entre 1990 et 2015. L'emploi de vingt-cinq termes identifiés comme faisant référence à la création d'informations par des particuliers est analysé. À partir de ce premier corpus composé d'articles scientifiques, les auteurs constatent, notamment, que les termes « *User-generated Content* »[46] et « science citoyenne » sont fréquemment employés, et que leur usage est croissant depuis les années 2010. Le terme de *crowdsourcing*, apparu dans le corpus en 2005, a connu une forte croissance en 2012 et a atteint plus de 1200 citations en 2015, tendance qui coïncide avec la croissance du procédé que nous développons dans le quatrième chapitre. C'est la plus forte occurrence parmi les vingt-cinq termes.

Dans un second corpus, les dix-sept chercheurs identifient une centaine d'applications et de sites web comme des cadres de production d'informations volontaires spatialisées. Par l'étude de ce second corpus, composé exclusivement de projets en ligne, il apparaît que le domaine de l'écologie provoque le plus d'intérêt parmi les communautés de contributeurs. Ce grand nombre de projets volontaires n'est pas surprenant compte tenu de la très longue histoire de la science citoyenne dans le domaine de l'écologie pratiquée par les sociétés savantes dès le XVIIIème siècle. Dans cette thématique de l'écologie, l'environnement était fortement représenté dans le corpus du fait de l'existence ancienne de stations météorologiques amateurs, de relevés de hauteurs de cours d'eau ou encore de systèmes de signalisation de catastrophes naturelles. Au contraire des projets spécialisés qui sont majoritaires, OpenStreetMap est signalé comme relevant d'un cadre de production diversifié. Il couvre des domaines très variés, aussi bien, par exemple ceux des transports que ceux concernant l'occupation du sol. Ainsi, l'information produite est-elle précisément géographique, au sens où les contributeurs dessinent des éléments cohérents topologiquement mais ont également la liberté d'exprimer les spécificités des territoires localement. Parmi ces cadres de production, les auteurs distinguent alors deux catégories d'appellations. La première

[46] Contenu généré par les utilisateurs

catégorie met l'accent sur la production volontaire d'informations par des particuliers comme dans l'*Information Géographique Volontaire* ou dans le *Contenu Généré par l'Utilisateur*. Le second type d'expressions souligne le processus ou les mécanismes mis en œuvre pour générer l'information. Les expressions comme « science citoyenne » ou « *crowdsourcing* » en sont des exemples.

L'appellation générique de « *crowdsourcing* » réunit les termes déjà cités. Le terme a été popularisé en 2006 par Howe[47], qui l'a utilisé pour décrire tout projet d'appel aux volontaires, qu'il soit étatique, associatif ou entrepreneurial. Il faut noter que Howe limitait ce cadre au contexte d'Internet. Le terme a été précisé par Enrique Estellés-Arolas et Fernando González-Ladrón-de-Guevara(2012)[48]. Ces deux auteurs définissent plus précisément le « *crowdsourcing* » comme l'externalisation de tâches par un organisme, le plus souvent commercial, vers « la foule ». Un organisme fait appel à un groupe d'individus et propose des tâches qui requièrent des compétences plus ou moins spécialisées à un nombre d'individus plus ou moins important. L'externalisation est un échange qui repose sur une rétribution mutuelle. Dans cet échange, les individus apportent leurs ressources et produisent selon la demande de l'organisme. La rétribution peut être économique. Elle est cependant souvent d'ordre moral (Estellés-Arolas et González-Ladrón-de-Guevara, 2012). Michael Goodchild (2012) rappelle que le terme prend sa source dans l'expression « *out-sourcing* » associé ensuite au terme désignant le groupe « *crowd* », foule, par Surowiecki (2005). Dans son livre, *The Wisdom of crowds*, Surowiecki soutient l'idée qu'une proposition issue d'un groupe en réponse à un problème sera plus utile et adaptée que si le problème n'était posé qu'à un unique individu, isolé bien que compétent. En effet, ce groupe peut, selon Surowiecki, ne pas être composé d'individus experts de la question pour être cependant en mesure de dépasser en qualité la réponse d'un expert. En complément de cette capacité d'un groupe à proposer des solutions originales, Goodchild (, 2012) souligne l'importance du nombre de producteurs qui composent ce groupe. Plus le nombre de

[47] D'après See et *al.*, 2016.

[48] « *Crowdsourcing is a type of participative online activity in which an individual, an institution, a non-profit organization, or company proposes to a group of individuals of varying knowledge, heterogeneity, and number, via a flexible open call, the voluntary undertaking of a task. The undertaking of the task, of variable complexity and modularity, and in which the crowd should participate bringing their work, money, knowledge and/or experience, always entails mutual benefit. The user will receive the satisfaction of a given type of need, be it economic, social recognition, self-esteem, or the development of individual skills, while the crowdsourcer will obtain and utilize to their advantage what the user has brought to the venture, whose form will depend on the type of activity undertaken* ». Fernando González-Ladrón-de-Guevara (2012)

participants est important, plus le groupe valide statistiquement des affirmations proposées par d'autres contributeurs. De même, plus une assertion sera récurrente, plus elle gagnera en crédibilité. La dimension du groupe est donc un critère qui participe de l'évaluation de la confiance qui peut être accordée à la production réalisée par le groupe. Cette évaluation apporte une garantie de crédibilité[49]. Ceci permet de comprendre l'importance du rôle d'Internet dans un tel cadre de production. Internet joue ainsi un rôle facilitateur pour transmettre la demande et pour connecter le plus grand nombre à un projet.

Le même phénomène de confirmation par le nombre s'applique à l'identification et à la correction d'erreurs. La loi Linus, proposée suite à l'observation du fonctionnement du mode de développement en *open source* (Raymond, 2001), a été reprise pour l'étude des productions d'informations en groupe. Ainsi, pour qualifier les productions du *crowdsourcing*, les chercheurs s'appuient sur ce qu'il est convenu de nommer la loi Linus : « Avec suffisamment d'yeux, tous les bugs sont superficiels » (Raymond, 2001, p.19). Dans son essai, l'auteur explicite les caractéristiques nouvelles du développement communautaire en *open source*. Le principe est qu'un grand nombre de contributeurs aux caractéristiques différentes permet le signalement rapide d'un problème, auquel un ou plusieurs membres proposeront une solution. Selon l'ampleur de la proposition de solution, cette dernière pourra soit être immédiatement mise en place, soit être mise en débat. Sur ce point précis, certaines différences distinguent les projets *open source* du projet OSM. Les travaux de développement d'un programme *open source* communautaire s'organisent généralement autour d'une équipe pivot. Celle-ci est appelée la *Core Team,* est chargée de garantir les fonctionnalités techniques pour lesquelles le projet est bâti. Les membres de cette équipe ont donc un pouvoir décisionnel sur les correctifs proposés ; cela leur permet parfois de les refuser. À l'inverse, dans OSM, ce qu'on pourrait appeler une *Core Team*, existe mais l'exercice de ce pourvoir est rare[50]. Si donc la loi Linus s'applique bien à OSM, c'est avec quelques restrictions.

En 2007, dans sa première définition de l'information géographique volontaire, Goodchild compare le VGI au concept, plus large, de « l'humain capteur »[51]. Il rappelle

[49] L'idée de crédibilité est précisée dans le chapitre 2.
[50] L'organisation sociale du projet OSM est analysée dans la sous partie 2.2 du chapitre 4.
[51] *Human sensors.*

que l'usage de réseaux de capteurs statiques ou portés passivement par des animaux ou par des humains, n'est pas nouveau. Cependant, dans le cadre du processus de création de bases de données d'informations volontaires, ce sont les réseaux de capteurs humains actifs, considérés comme des individus capables d'analyser et d'interpréter ce qu'ils observent, qui s'en rapprochent davantage. Il précise deux cadres dans lesquels ce mode de production pourrait être mobilisé. Évoquons d'abord certains domaines scientifiques qui s'appuient sur des communautés d'amateurs. Leur rôle d'observateurs[52] dans un projet de recherche apporte la couverture nécessaire d'un territoire pour des analyses à petite échelle. L'exemple célèbre du *Chrismas Bird Count* est mobilisé par Goodchild pour illustrer son propos. Il s'agit d'individus connaisseurs en ornithologie, mais qui ne sont pas des professionnels, auxquels l'association a transmis des protocoles de collectes, assurant ainsi une production homogène d'informations entre chaque observateur. Goodchild considérait le VGI comme une utilisation efficace de ce réseau, rendue possible par le Web 2.0[53]. Si on le compare aux projets contributifs sur Internet, on identifie une différence notable : les participants aux projets contributifs subissent moins de contraintes. Ils collectent des connaissances en apportant parfois des compétences élevées, sans pour autant être tenus de suivre un protocole conçu à l'avance. Certaines associations ont la possibilité de former les participants à un protocole et de les guider vers une production, suivant des modes opératoires proches des productions traditionnelles. A l'inverse le *crowdsourcing* repose, sur des contributions nombreuses mais moins protocolaires. L'information géographique volontaire est à placer dans ce second cadre puisqu'il était encore flou au moment où Goodchild l'évoquait en 2007.

Le deuxième cadre, abordé en 2007 par Michael Goodchild, est un cadre qui mobilise des humains-capteurs sans aucune attente et qui accepte la participation de néophytes. À titre d'exemple, Michael Goodchild évoque la signalisation de la création d'une nouvelle rue par des acteurs locaux auprès d'organismes de cartographie. Ce type d'information n'exige aucune connaissance de protocoles, ni de méthodes de collecte. L'opérateur, qui réceptionne des informations simples, est l'acteur de la création de l'information. En outre, on peut intégrer à ce cadre de production volontaire les participants quasi passifs qui activent un capteur et transmettent l'information sans

[52] *Citizen science*.

[53] « *Humans as sensors One can see VGI as an effective use of this network, enabled by Web 2.0 and the technology of broadband communication* » (Goodchild, 2007).

aucune participation au processus de création, en dehors du fait d'avoir activé le capteur.

En 2016, See et *al*. présentent également des classifications de contextes de production d'informations qui suivent le modèle du *crowdsourcing* (fig. 5). Ils précisent que les productions issues des particuliers peuvent être passives (involontaires ou volontaires) ou actives (volontaires). Les contextes de production mobilisant des contributeurs actifs présentent alors des degrés différents d'implication de la part des contributeurs. Les vingt-cinq termes identifiés par See et *al*. dans leur premier corpus, sont organisés selon une dichotomie qui oppose les créations passives et les créations actives. Seuls deux termes évoquent une production involontaire. La première notion est « l'information géographique contribuée »[54] (Harvey, 2013) qui sous-entend que le producteur n'a pas nécessairement connaissance qu'une information est collectée. La seconde est « l'information géographique involontaire » (iVGI)[55] qui désigne une absence totale d'action de la part du producteur, qui ignore même qu'une information est collectée. De manière moins claire, « l'information géographique contextuelle »[56] (AGI) désigne un mode de production qui ne prévoit pas de participer à une base de données géographique mais qui apporte des informations contextuelles, indirectement par la publication, de photographies, de messages ou de vidéos. En 2013, l'expression AGI est proposée (Stefanidis et *al*., 2013) pour désigner les empreintes géographiques laissées par les participants aux média sociaux. Les informations qui peuvent être collectées par les média sociaux sont alors essentiellement des localisations. Ainsi, en associant une localisation à leurs publications, sur *Twitter, Flickr* ou *Youtube*, les auteurs participent-ils largement à la création de ressources en informations spatiales sur des individus. La quantité d'informations disponible et les analyses géographiques réalisées sur ces informations ont permis la production d'informations géographiques à partir de ces données spatiales. Des analyses géographiques de ces données peuvent mettre en évidence des effets de centralité. Comme pour « l'Information géographique contribuée » (CGI), les AGI sont des données peu spécialisées et non conceptualisées. Mais elles sont produites en grande quantité.

[54] « *Contributed Geographic Information* » (Harvey, 2013).
[55] « *Involuntary geographic information* ».
[56] « *Ambient Geographic Information* ».

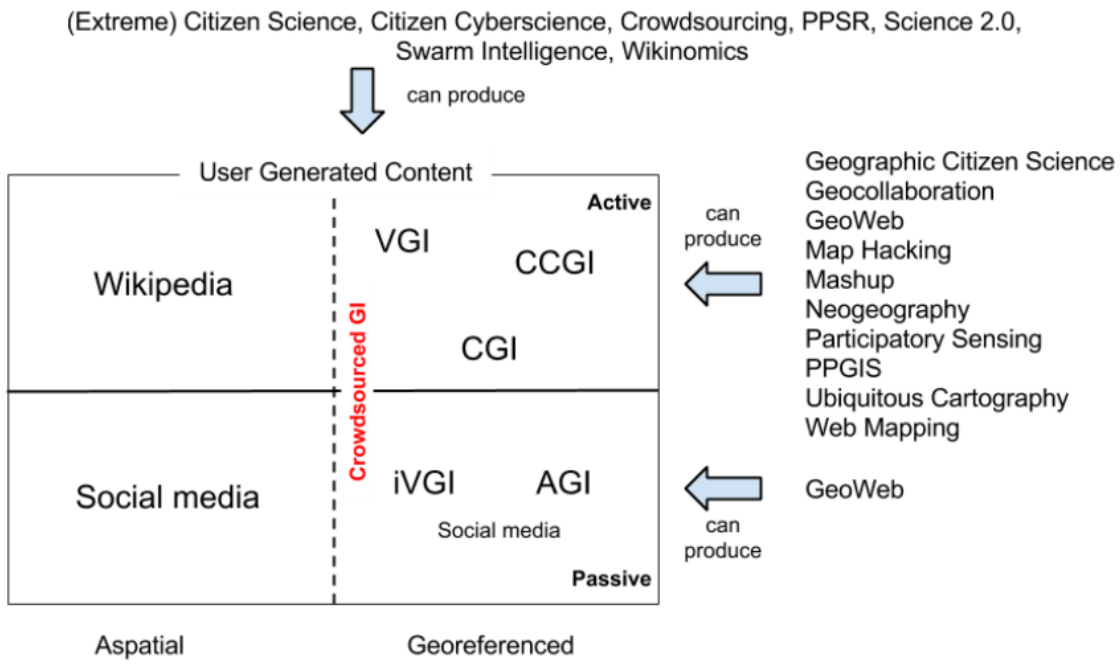

Figure 5 See et al., 2016 ; synthèse des expressions relatives à la création d'information par des particuliers

Un deuxième ensemble de termes est mis en évidence à partir du premier corpus cité précédemment. Ils désignent des contextes de production permettant exclusivement une production active, et donc volontaire (See, *al.*, 2016) : « science citoyenne », « science citoyenne extrême », « information géographique contribuée par les citoyens »[57], « cyber-science citoyenne », « *crowdsourcing* », « participation publique à la recherche scientifique »[58] (PPSR), « Science 2.0 », « intelligence distribuée »[59], « wikinomie »[60] et « information géographique volontaire ». L'expression « science citoyenne » date des années 1990 et n'est pas directement liée au cadre d'Internet (See et *al.*, 2016). Elle désigne la proposition, par le milieu académique, d'un cadre permettant au grand public de participer à des projets de recherche. Cette production repose sur une participation intellectuelle active du grand public qui y apporte ses ressources en connaissances et sa main d'œuvre.

L'analyse du 2ème corpus concernant les sites Web et applications (See et *al.*, 2016), montre que la majorité des projets sont des cadres de productions actifs. Dans le domaine de l'information géographique, la « science citoyenne en géographie »[61], la

[57] « *Citizen-contributed Geographic Information* ».
[58] « *Public Participation in Scientific Research* ».
[59] « *Swarm Science* ».
[60] « *Wikinomics* ».
[61] « *geographic citizen science* ».

« Geo-collaboration », la pratique du *hacking* appliquée à la cartographie[62], le *mashup*, la néogéographie, la « participation publique aux systèmes d'information géographique »[63](PPGIS) et la cartographie en ligne[64] sont des variantes de ces contextes de production actifs (See et *al*., 2016).

Parmi les cadres de productions actifs, plusieurs degrés d'implication des contributeurs ont été identifié (Haklay, 2013 ; Sui et *al*., 2013). Associés à cette distinction, certains contextes de production mobilisant des contributeurs actifs, présentent des opportunités de création plus ou moins complexes. De même, les cadres de contribution actifs, présents dans le corpus de See et *al*., ont mis en évidence qu'un nombre limité de termes couvre des thématiques dont les institutions peuvent être responsables. Par exemple, l'enregistrement d'événements comme des incendies, la mise en valeur de lieux touristiques, le recensement d'espèces ou la collecte locale de phénomènes naturels sont des thématiques qui peuvent être institutionnelles mais qui ne sont pas partagés par de nombreux pays. En outre, ce ne sont pas des thématiques pour lesquelles un institut national disposerait de larges investissements nationaux. Les projets contributifs relatifs aux thématiques couvertes par des instituts nationaux sont plus rares (Wikimapia et OpenStreetMap). Selon See et *al*., cette distinction entre occupation des groupes de *crowdsourcing* et les institutions s'explique par un écart d'expertise entre les groupes. D'après leur corpus de projets de *crowdsourcing*, les compétences les plus élevées qui sont pratiquées relèvent le plus souvent de la collecte de coordonnées. En outre, l'ergonomie des sites de collecte par *crowdsourcing* facilite la participation, en exigeant des contributeurs qu'ils se connectent à Internet et remplissent des formulaires. Cependant, les auteurs soulignent une gradation dans les compétences nécessaires pour participer aux différents projets. Il est intéressant de noter que les compétences ne sont pas attendues *a priori* des contributeurs. Elles sont transmises par les tutoriels, écrits ou vidéos, par des séances d'entraînements et par des documentations. Des tâches plus ou moins expertes, dans un même projet contributif, permettent ainsi aux participants néophytes de se familiariser avec les spécificités du contexte de production auquel ils s'intègrent.

[62] « *Map Hacking* ».
[63] « *Public Participation in Geographic Information Systems* ».
[64] « *Web mapping* ».

Face à ces participations non rémunérées, la question de la motivation des contributeurs apparaît. En effet, rien n'assure la persistance de la motivation chez les contributeurs. Cette absence de garantie affaiblit, *a priori*, la fiabilité des projets contributifs. La littérature scientifique identifie deux mécanismes de motivation : l'engagement au nom de valeurs de partage et l'obtention d'une reconnaissance de compétence au sein d'un groupe d'experts. Dans le cadre du projet de recherche ECCE Carto (des Espaces de la Contribution à la Contribution sur l'Espace), mené au sein du laboratoire PASSAGES (UMR5319), une enquête socio-géographique auprès des contributeurs français a été conduite au cours de l'hiver 2015-2016. Comme les résultats de l'enquête le précisent, ce sont les contributeurs les plus « connectés » aux réseaux sociaux numériques et physiques qui ont répondu à l'enquête (298 enquêtés). « Cette enquête dresse donc un panorama d'une cible particulière : les contributeurs actifs et connectés à la communauté OSM ». Les pratiques de contribution, les valeurs d'engagement et les profils socio-économiques des contributeurs français sont analysés. Ainsi, la volonté de participer à une production selon un mode collaboratif motive-t-elle 20 % des enquêtés. L'accès gratuit, la simplicité de réutilisation et le dynamisme communautaire sont perçus comme des motifs attractifs. Enfin, l'importance du lien entre le projet OpenStreetMap et le courant du Libre y est souligné (Duféal et *al*., 2016). Dans une perspective qualitative, Budhathoki et Haythornthwaite (2013) se sont appuyés sur la littérature scientifique relative aux motivations de participation aux projets de développement de codes informatiques et de *wikis*, pour proposer une étude de cas propre à OpenStreetMap. En décembre 2009, Budhathoki et Haythornthwaite ont diffusé un appel à témoignages, au moyen de la messagerie d'openstreetmap.org, à destination de contributeurs identifiés comme actifs par l'analyse de leurs contributions. Sur les 31000 contributeurs contactés, 459 ont répondu à l'enquête (moins de 2%). L'enquête visant à tester des hypothèses de motivation a permis de mettre en évidence une logique de motivation « positive » dite « de besoins personnels mais partagés »[65]. En d'autres termes, les contributeurs servent leurs intérêts personnels à l'échelle individuelle, tout en contribuant à servir les intérêts d'autres individus. Ces intérêts sont communs mais ils diffèrent très clairement entre les contributeurs très actifs, dans le projet et dans l'animation de la communauté, et les contributeurs plus occasionnels (Haythornthwaite, 2009). Les premiers tirent de leur activité, une

[65] « *Personal but shared need* ».

valorisation professionnelle mais surtout une renommée au sein de la communauté, par leur rôle de médiateurs ou d'animateurs sur les *wiki*s et les forums. Par cette communication, les contributeurs très actifs soutiennent la pérennité du projet auquel ils participent en assurant la formation des autres contributeurs et la clarification des règles gouvernant le projet. Une reconnaissance de leurs compétences techniques est acquise par ce biais ainsi qu'une reconnaissance de leur connaissance du fonctionnement du système global du cadre de production. Les contributeurs occasionnels semblent davantage orientés par le soutien des principes du Libre, de la réutilisation, souvent plutôt non-commerciales, des données et la mise en valeur d'un territoire par la diffusion de connaissances locales (Budhathoki et Haythornthwaite, 2013). L'idée d'une récompense est également à associer à celle de la motivation. D'après les deux enquêtes citées précédemment, la constitution d'un réseau social est la principale récompense. Ce réseau social reconnaît la participation du contributeur et lui accorde, par-là, une reconnaissance sociale.

Le mode de production d'informations volontaires repose sur les technologies du web, sur des mécanismes de reconnaissance au sein d'un groupe, qui doit être composé d'un certain nombre de contributeurs, assurant le développement de la ressource au cœur du projet contributif. Le partage d'informations mais également l'accès de tous à l'édition de ces informations est une des motivations principales de la participation à un projet contributif. Cependant, des études sur les profils socio-économiques des contributeurs à OpenStreetMap montrent que les populations les plus impliquées dans OpenStreetMap sont de sexe masculin, hautement éduqués, résident en zone urbaine dense et sont compétentes, professionnellement, dans les métiers de l'informatique, des SIG ou des traitements de données. Les résultats de l'enquête, réalisée en France, dans le cadre du projet ECCE Carto, vont dans ce sens. Sur les 298 enquêtés, 88 % sont des hommes, d'âge moyen 38 ans et 58% des répondants ont au moins un niveau Master. Enfin, 60 % des enquêtés ont eu accès, durant leurs études ou leur parcours professionnels, à une formation informatique. Arsanjani et *al*. (2015) rapportent des résultats similaires en prenant pour individus des cellules de 1000m² d'une grille couvrant le land de Bade-Wurtemberg. Leurs résultats montrent une corrélation nette entre des zones densément éditées par les contributeurs à OpenStreetMap et des caractéristiques socio-économiques de territoires favorisés. Ainsi, les zones à forte densité, dont la population présente un niveau d'éducation moyen, des revenus élevés,

un nombre élevé d'étrangers et une population résidente majoritairement située dans la tranche d'âge des 18 à 69 ans sont les zones les plus susceptibles d'être fortement documentées dans OpenStreetMap.

L'ampleur du projet contributif OSM en fait un objet de recherche à plusieurs facettes. OSM est un objet de recherche en tant que source de données potentielles à évaluer. Il s'agit également d'un cadre de production qui s'inscrit dans la durée (quinze ans à la date de rédaction de ce texte) tout en étant original. Ainsi, le mode de production conçu par les contributeurs, mais également désormais par des entreprises utilisatrices d'OSM, est un sujet de recherche pour les géographes, pour les cartographes et pour les chercheurs en sciences de l'information et de la communication. L'étude de chacune de ces facettes nécessite de comprendre les motifs de la création du projet, la structuration de son objet principal, c'est-à-dire la base de données ainsi que les principales modalités de contribution. La section suivante en fait son objet principal.

1.2.3. OpenStreetMap : une base de données géographique contributive

En 2004, Steve Coast, étudiant en informatique, est à la recherche de données géographiques gratuites sur la ville de Londres, afin de poursuivre des recherches en géographie. À l'époque, de telles données ne sont diffusées que par l'*Ordnance Survey* et sont payantes. Il a alors l'idée d'élaborer, à l'*University College London* (UCL), le projet d'une carte mondiale ouverte (OpenStreetMap) à tous et ouverte sur le Monde. Publier des données ouvertes conçues avec des logiciels ouverts sont ses motivations initiales. Les choix qui en découlent reposent alors sur les principes du Libre tout comme le choix de n'imposer aucune ontologie aux contributeurs[66]. Dès les premières communications publiques, Steve Coast assure que le projet OpenStreetMap (OSM) n'est pas seulement celui de la production d'une carte des rues, qui serait utilisable par tous mais précisément celui d'une base de données relative aux éléments physiques de l'espace public, dont le mode de collecte repose essentiellement sur la technologie du GPS et se trouve documenté par des métadonnées. Ainsi, le cœur du projet est une base

[66] Steve Coast présente en mai 2005 les principes initiaux de son projet, encore d'actualité en 2019 : https://blog.openstreetmap.org/wp-content/uploads/2005/05/osmtalk.pdf

de données librement éditable et en accès libre, selon les conditions d'une licence de type *copyleft*[67]. Le travail concentré sur la base de données doit aussi ouvrir le champ des usages aux calculs d'itinéraires, au guidage, à la cartographie thématique, aux *mashups* ou encore au géocodage.

La base de données OSM est constituée d'objets géographiques de l'espace public. Seuls des phénomènes visibles et délimités sur le terrain (côtes, forêts, bâtiments, réseaux routiers) ou formellement admis (découpage administratif) sont cartographiés. Cette convention assure la possibilité, pour d'autres contributeurs, de confronter l'objet réel et sa description dans la base de données. Ils pourront alors être amenés à corriger ou à enrichir l'objet enregistré dans la base de données. Aucune information non observable n'est acceptée dans l'écriture des informations. C'est là une des rares limites à la création *de tags* (étiquettes). Les objets sont conçus pour des représentations à grande échelle avec pour objectif l'exactitude géographique du tracé. Chaque objet de la base principale a pour élément (fig. 6), soit un nœud (*node*), soit un chemin (*way*) soit une relation (*relation*). Chacun de ces objets (à l'exception des nœuds créés pour la construction de chemin) sont *tagués* de manière à les catégoriser et à construire des pratiques de contribution communes. Les *tags* d'OpenStreetMap sont ainsi conçus comme des couples `clé=valeur`. La pratique de la contribution dans OpenStreetMap relève donc de la participation à une folksonomie[68] en respectant des conventions. Les objets de la base sont classés spontanément par les contributeurs sans contrainte formelle, en principe.

[67] La définition précise de cette licence est fournie dans la section 1.2.3.

[68] Rappelons que le terme folksonomie désigne la pratique de classification d'informations (photographies, pages web, signets) par les utilisateurs d'Internet qui marquent les ressources disponibles sur le réseau par des étiquettes (ou *tags*). La folksonomie est définie par les auteurs, s'appuyant sur Wal T.V. in « Folksonomy Definition and Wikipedia » (2005) comme « le résultat de l'activité d'association libre d'étiquettes (*tags*) à des données ou objets (tout ce qui est identifié par une URL) dans le but personnel de les retrouver ultérieurement » (Hombiat et *al.*, 2015, p.106).

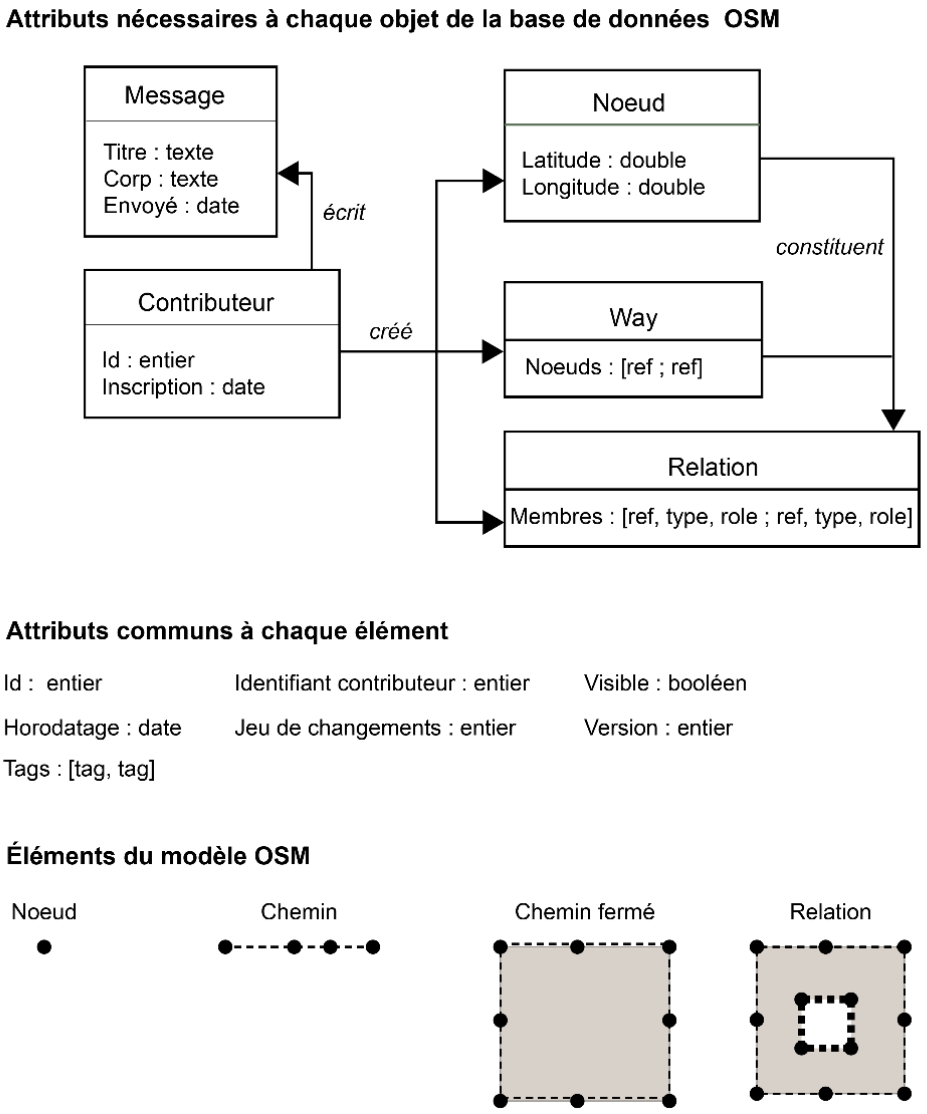

Figure 6 Éléments et métadonnées de la base de données OpenStreetMap (Hayat, 2019)

Au moment de la création d'un objet, le contributeur doit, dans un premier temps, choisir à quel élément (interprété comme un objet géométrique lorsqu'il est chargé dans une base de données géographique) il associera des *tags* (attributs thématiques en base de données relationnelle). Pour qu'un objet soit ajouté à la base, sept informations sont nécessaires : les coordonnées de l'objet, l'identifiant de l'objet, l'identifiant du contributeur, si l'objet est visible ou non, la version de l'objet, la date de création ou de modification et le jeu de changement (ou *changesets*) durant lequel l'objet a été modifié. Ces informations (à l'exception des coordonnées) sont créées automatiquement lors de l'édition de la base par un contributeur. Ces informations sont les métadonnées des objets. Réunies, ces métadonnées renseignent sur le contexte de production de l'information. Souvent, le logiciel par lequel l'objet a été créé est

également renseigné. Paradoxalement, les informations contribuées dans la base de données centrale du projet OSM ne sont pas stockées selon un format numérique spécifique à des données géographiques. En effet, les copies brutes de la base de données OpenStreetMap sont diffusées sous les formats XML et PBF. Chacune des copies est une base de données contenant l'information uniquement sous la forme de nœuds. Chaque nœud porte des coordonnées géographiques, ainsi que l'information selon laquelle il est partie d'un chemin ou s'il est isolé, on peut aussi ajouter éventuellement des tags. C'est au chargement d'une copie en une base de données que les objets sont écrits de manière à s'intégrer dans des chaînes de traitement de données géographiques.

Les objets sont donc construits d'abord selon un élément (ou géométrie), puis des caractéristiques sémantiques (*les tags*) peuvent venir les qualifier. Dans un second temps, le processus d'itération de la contribution est sensé mener à la précision de la catégorisation de l'objet par des *tags* plus spécifiques. Nous pouvons citer, par exemple, quelques objets conçus comme des ponctuels :

- une source d'eau naturelle (`natural=spring`),
- le centre d'une localité (`role=admin_centre`) ou encore
- une sortie d'autoroute (`highway=motorway_junction`),

Ces objets sont conçus en nœuds du fait de leur taille (une source), de leur caractère conventionnel et non physique (le centre d'une zone administrative, souvent placé devant le bâtiment représentant l'autorité locale) ou encore afin de contribuer à la construction d'un graphe dont le nœud est un sommet (une sortie d'autoroute connectant une autoroute et une route du réseau inférieur).

Les linéaires (chemins ou *ways* ouverts) et les surfaces (chemins fermés) sont composés de nœuds ordonnés. Par le dessin d'un chemin ouvert seront évoquées une route principale (`highway=primary`), une rangée d'arbres (`natural=tree_row`). Un chemin fermé permet la définition de zones comme des bâtiments (`building=*`) ou des zones officielles (`boundary=administrative`). Lorsque l'information est précisée, une route est caractérisée par un nombre de voies (`lanes=*`), un revêtement (`surface=*`), une vitesse de circulation maximum (`maxspeed=*`), mais ne peut être qualifiée d'attractive du point de vue du seul contributeur. C'est la mention de « points de vue » (`tourism=viewpoint`) ou d'éléments physiques, comme un sommet de

montagne (`natural=peak`) qui permettront de signaler l'attrait de ce paysage, dans le cas de l'utilisation des données par la cartographie touristique. Cependant, si un panneau indique une reconnaissance officielle de l'attrait d'une route, cette information pourra être mentionnée. Par exemple, un lieu classé Unesco (`heritage=1` et `heritage : operator=whc`) est une information dite objective.

Les contributeurs sont libres de choisir quels *tags* seront les plus appropriés pour retranscrire leurs connaissances. En pratique, la diversité des *tags* ainsi que certaines divergences au sein de la folksonomie (fig. 7) mènent nécessairement à une forme de complexité. L'utilisation de la base de données OpenStreetMap pour la cartographie requiert donc une généralisation parfois géométrique, selon l'échelle de la carte, mais surtout thématique. Par exemple, dans le cadre d'une carte touristique, le cartographe peut souhaiter représenter par un symbole unique les espaces verts qu'il aura sélectionnés. Le cartographe pourra être amené à réunir sous ce symbole les *tags* suivants : `leisure=garden`, `leisure=park` et à ne conserver parmi les jardins que ceux de type botanique : `garden_type=botanical`.

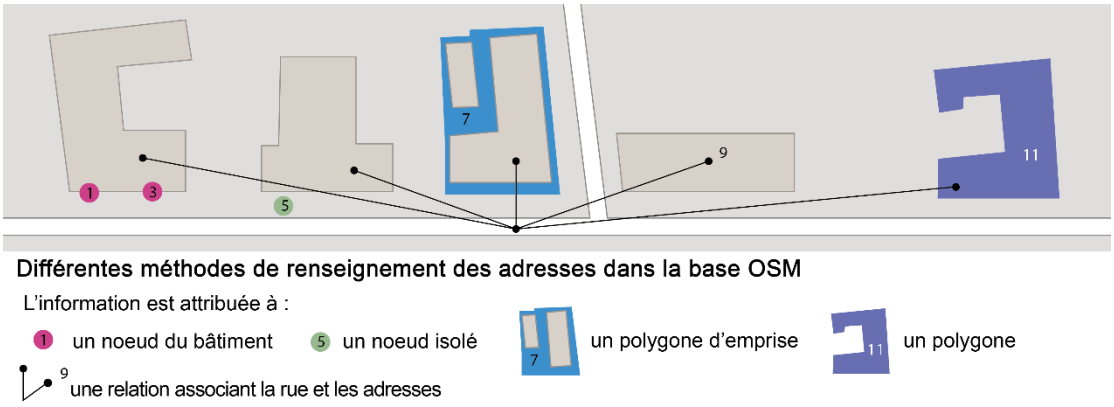

Figure 7 Différentes méthodes de contributions des adresses (Hayat, 2019)

Dans OpenStreetMap, l'aspect chaotique qui accompagne parfois la catégorisation d'informations par la folksonomie est empêché par un partage méthodique du bon usage des *tags*. En effet, la contribution commence souvent par un passage en revue d'articles du *wiki* OpenStreetMap[69]. Des conseils techniques et thématiques sont disponibles sur le site. En s'appuyant sur la documentation que constitue le *wiki*, le contributeur peut

[69] Map Features (2019)

choisir d'utiliser les *tags* couramment employés[70] ou d'en proposer de nouveaux, qui correspondent davantage à la situation qu'il connaît.

Si la libre contribution est un facteur de complexité, elle repose cependant sur un socle commun : contribuer suppose de connaître les choix de catégorisation des précédents contributeurs et de suivre des conventions (uniquement des données objectives et une géométrie faiblement généralisée). Le principe du Libre, très présent dans le choix d'une classification en folksonomie, exclut les contraintes strictes. Cependant, progressivement les bonnes pratiques écrites dans le *wiki* ou simplement généralisées dans la pratique s'approchent du procédé classique de définition de spécification pour une classification. Ainsi Hall et *al*. (2018) se sont demandé dans quelle mesure la documentation écrite par les contributeurs peut être désignée comme un standard. L'étude menée sur des données, extraites en février 2014, prend pour emprise les frontières des États-Unis. Une série de *tags* sont sélectionnés dans le but d'étudier les objets décrivant des commerces de chaînes alimentaires comme les McDonnald's ou les 7-Eleven. L'expérience consiste en un test d'homogénéité de description entre les magasins de ces chaînes. La cohérence des *tags* en fonction des recommandations du *wiki* en tant que standard d'écriture des données est également testée. L'étude démontre que la contribution est très majoritairement cohérente avec les instructions du *wiki*. Néanmoins, les contributeurs tendent à décrire les éléments généraux tels qu'il est recommandé dans la documentation (cet objet représente un café) mais négligent les *tags* secondaires pourtant documentés dans le *wiki* (les horaires d'ouverture, le type de cuisine préparé). Le *wiki* OpenStreetMap est donc un « standard informel » (Hall et *al*., 2018) partiellement respecté mais rarement enfreint.

La base de données OSM est le produit le plus évident réalisé par les contributeurs. Néanmoins, le projet OSM a pour visée la construction d'un système complet, permettant le partage libre d'informations géographiques. De nouveau, ce sont les principes du Libre qui imprègnent les directions choisies par les contributeurs. Trois grandes catégories de pratique d'OSM sont distinguées dans la page d'accueil du *wiki*, recueil des principes, des méthodes et des débats choisis par les contributeurs :

- *l'utilisation* d'OpenStreetMap,

[70] Topf, J. The TagInfo Webservice—Statistics About Tags in The OpenStreetMap Database. URL : http://taginfo.openstreetmap.org/keys

- *la contribution* au développement de la base de données
- et *le développement* logiciel.

L'utilisation répond au principe de partage des données sans restriction d'accès ; *la contribution* est ouverte à qui le souhaite, afin de bénéficier des connaissances de chacun par itérations ; *le développement* logiciel correspond à la pratique majoritaire du domaine du Libre et soutient l'objectif d'indépendance face à des logiciels propriétaires, dont les clauses juridiques ne correspondraient pas aux exigences de transparence et de confidentialité du Libre. Ainsi, afin de répondre à l'objectif du projet contributif, des contributeurs ont-ils développé progressivement une infrastructure technique originale (fig. 8).

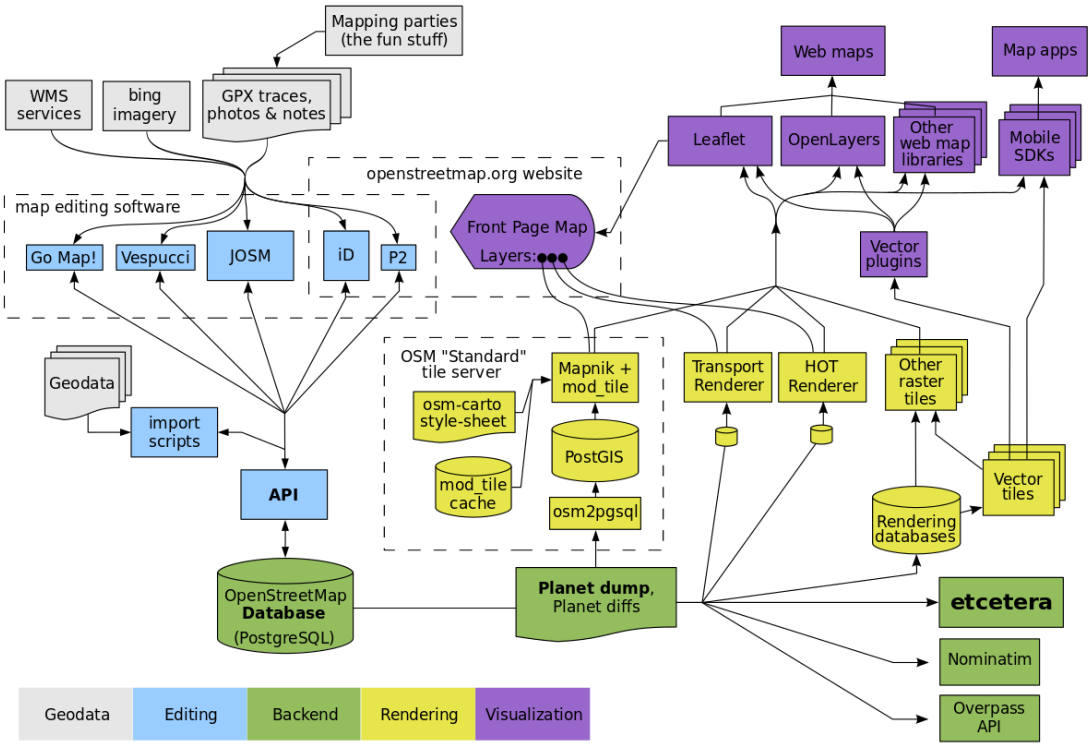

Figure 8 Synthèse des composants de l'infrastructure actuelle permettant le maintien et l'utilisation de la base de données OSM[71]

L'élément central de cette infrastructure est la base de données OpenStreetMap actuellement enregistrée selon les règles du système de gestion de base de données (SGBD) PostgreSQL. Ce composant technique, qui compte parmi les figures du Libre, publié et perfectionné depuis 1995, remplace depuis 2009 le Système de Gestion de

[71] Image réalisée par Harry Wood en juillet 2017, disponible sur le wiki : https://wiki.openstreetmap.org/wiki/File:OSM_Components.svg#filelinks

Base de Données *MySQL*[72]. La base de données OSM est conçue comme le point névralgique de toutes les contributions de données, elle est donc unique (mais néanmoins souvent dupliquée) et le seul moyen de l'éditer est d'utiliser l'interface de programmation applicative d'OSM (API OSM). L'API OSM est une interface de service en ligne spécifiquement développée pour permettre un fonctionnement contributif entre internautes qui ne sont pas nécessairement en contact. La version actuelle de l'API est la 0.6. La première version publiquement utilisée est la 0.3. Les trois versions publiques suivantes ont été remplacées entre mai 2007 et avril 2009, date à laquelle l'API 0.6 est mise en service (*API*, 2019). Si le passage de la version 0.3 à la version 0.4 a lieu pour des questions de développements internes, les deux autres changements de versions de l'API comptent comme des moments marquants du projet OSM.

En octobre 2007, l'API 0.4 est abandonnée au profit de l'API 0.5, qui permet la création de relations et le versionnement de chaque élément de la base de données. La notion de relation, propre au projet OSM, et l'enregistrement de chaque version ainsi que l'accès aux étapes d'itérations sont des premiers jalons dans la spécificité du projet contributif. La version 0.6 complète ces innovations en ajoutant de nouvelles contraintes aux éditions des données. Par exemple, un maximum de 2000 points par chemin est défini, dans l'optique de garantir l'utilisabilité de chaque élément. Nous retenons également la conception d'un ordre entre les éléments qui constituent une relation, dans le but de permettre la création de multi-polygones et de lignes de transport par exemple. Dans le cadre d'un projet contributif en ligne, des conflits d'ordre au moment de l'envoi des éditions vers la base de données centrale sont envisageables. Deux contributeurs peuvent éditer un même élément, au même moment. Afin d'empêcher une confusion au moment de l'écriture des nouvelles versions, une contrainte supplémentaire est prévue dans cette version de l'API sur le numéro de version. Enfin, un nouveau concept est créé, celui du *changeset* ou ensemble de modifications. Un *changeset* est ouvert lorsqu'un contributeur commence l'édition. Il peut le fermer à tout moment, après avoir commenté les modifications réalisées. Ce nouveau formalisme vise à davantage de communication et d'explicitation des modifications. Il est possible de préciser une source générale à toutes les modifications ou de préciser le contexte de production.

72 « OpenStreetMap moves to PostgreSQL », article du 28 avril 2009:
http://blog.cleverelephant.ca/2009/04/openstreetmap-moves-to-postgresql.html

L'API est donc le moyen technique permettant, via Internet, de lire la base de données et d'écrire dans la base de données.

Ces deux étapes, propres à la maintenance d'une base de données, sont réalisées via des éditeurs logiciels. La figure 8, représentation de l'infrastructure OSM disponible sur le *wiki* et régulièrement mise à jour, présente les principaux logiciels d'édition à la disposition des contributeurs pour modifier la base de données. Elle date ici de 2017 et présente cinq éditeurs. Parmi ces éditeurs, deux sont des interfaces en ligne, accessibles via openstreetmap.org (iD et Potlach 2 ou *P2*), deux sont des applications mobiles (Vespucci et GoMap!) et le dernier est un éditeur autonome (JOSM). La précédente représentation de l'infrastructure date de 2012 (figure 9) et ne présentait que trois éditeurs, deux autonomes (Merkaator et JOSM) et un en ligne (Potlach 2).

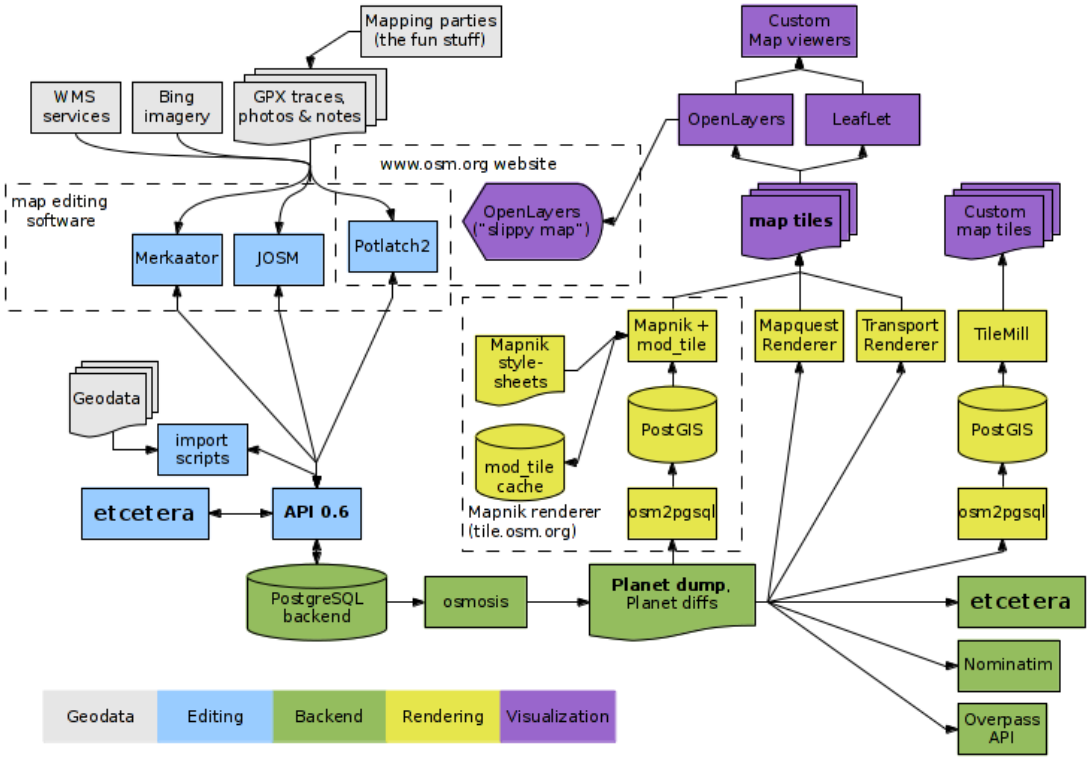

Figure 9 Synthèse datant de 2012 des composants de l'infrastructure permettant le maintien et l'utilisation de la base de données OSM[73]

Le logiciel JOSM est le plus prisé des contributeurs experts. Il permet de charger simplement des sources de données tierces et offrent de nombreuses fonctionnalités de

73 Image réalisée par le contributeur *Ivansanchez* en aout 2012, disponible sur le *wiki* : https://wiki.openstreetmap.org/wiki/File:OSM_Components.png

dessin et de fusion des objets. Il permet également d'intégrer des *plugins*[74] pour des besoins spécifiques comme le *plugin* Cadastre-fr qui permet, pour le territoire français, d'accéder aux images délivrées par la Direction Générale des Finances publiques (fig. 10). Cependant, le logiciel présente une interface compliquée et révèle toute la complexité de la création dans OSM. On comprend alors pourquoi deux applications mobiles et une application en ligne sont mises en valeur dans la représentation de 2017 (fig. 8). En outre, ces trois applications répondent à l'objectif d'ouverture de la pratique de l'édition vers un plus grand public.

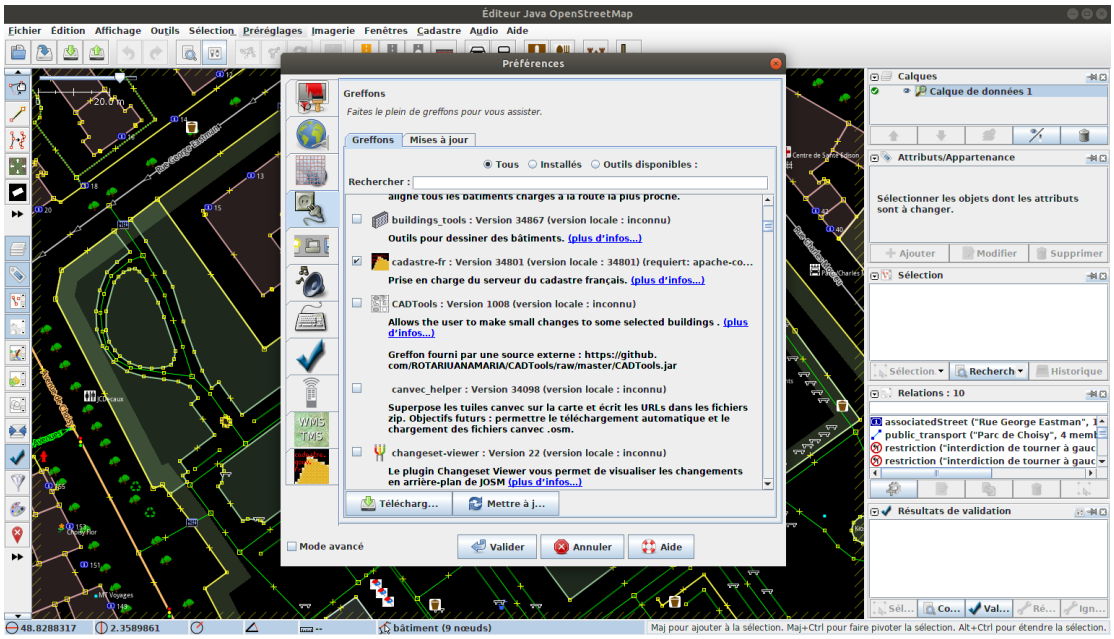

Figure 10 Extrait de la fenêtre d'installation des plugins dans JOSM

Les parties droites des figures 8 (p.71) et 9 (p.73) présentent les moyens conçus ou mobilisés par la communauté afin de créer des services à partir de la base de données. La première étape est la production de copies brutes (*dump*) de la base de données et de fichiers de différentiels (*diff*s). Ainsi, il est possible de constituer une base de données autonome de la base centrale constamment éditée et de maintenir la copie à jour par le moyen des différentiels. L'ensemble de l'infrastructure est conçu afin de permettre aux utilisateurs de mobiliser des outils libres comme osm2pgsql, afin de charger des données en format PostGIS. Il est possible, par la suite, de produire des cartes en ligne tuilées mais également des outils comme l'API Overpass ou l'outil de géocodage et de géocodage inversé Nominatim. L'API Overpass permet de lire la base

74 Un plugin est un ensemble logiciel répondant à une fonctionnalité spécifique et conçu pour être ajouté à un logiciel autonome.

de données et d'afficher les résultats d'une requête ; en cela elle diffère de l'API OSM. Par le moyen d'Overpass, il est possible de retrouver des objets qui partagent des *tags* communs, des objets créés à une date précise ou édité par un même contributeur. Le même principe de requête sert l'API du service TagInfo. Le service TagInfo a pour visée la production de statistiques relatives aux *tags* et, par extension, aux clés et aux valeurs. Il s'agit d'un outil automatique, complémentaire du *wiki* dans l'expression des pratiques de contributions. Par exemple, la figure 11 est une capture d'écran du résultat de la requête demandant les valeurs de la clé `amenity` par ordre d'utilisation. Les trois premières valeurs les plus utilisées sont « `parking` » (21%), `place_of_worship` (7,4%) et `bench` (7%). L'API compte 9313 valeurs différentes à la clé `amenity`. Cependant, passé un certain seuil, et selon la spécificité de la catégorie recherchée, on considérera les faibles pourcentages comme l'indice d'erreur de contribution. Par exemple le *tag* `amenity=market` est seulement compté 206 fois. Cela semble peu réaliste, compte tenu d'un type de pratique aussi répandu qu'un marché. En fait, le *tag* réellement choisi par les contributeurs est `amenity=marketplace` et il est compté 54046 fois en février 2019.

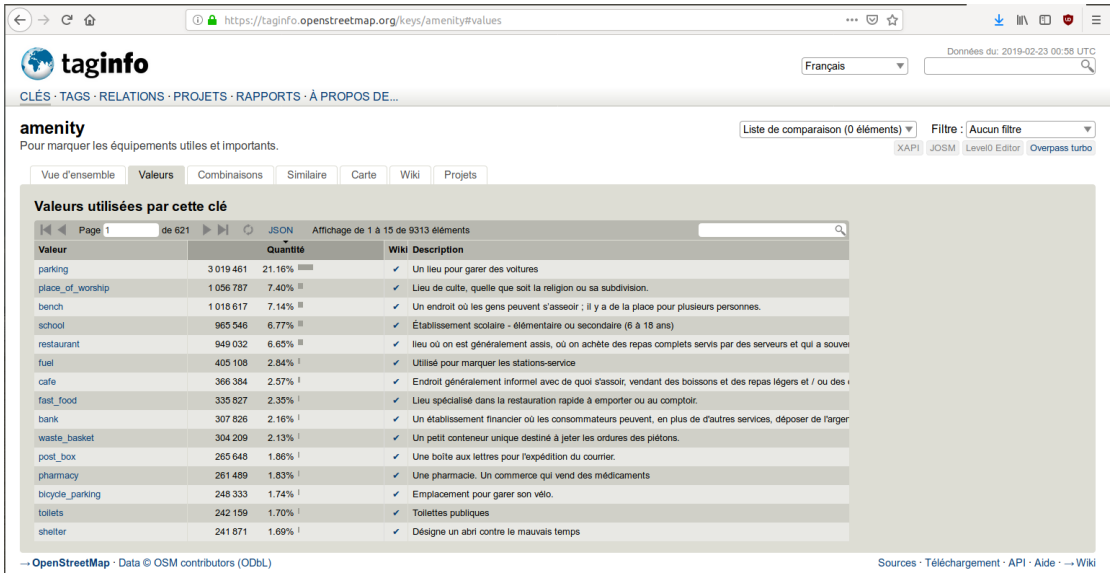

Figure 11 Les valeurs principales de la clé `amenity` en février 2019, capture d'écran du site taginfo

Le projet s'appuie désormais sur plus de dix ans d'expériences diversifiées des réalités du terrain. Les contributeurs décrivent librement les informations relatives à l'espace public. Afin d'exploiter une quantité importante d'informations contenues dans la base

de données, la pratique de contribution a mené à une organisation spécifique. À force de comparaisons, de généralisations et de discussions en communauté, les contributeurs sont parvenus à un socle commun de modélisation des espaces publics.

Pour autant, comme pour tout autre usage de la base de données, ces conventions ainsi que leurs variations doivent être maîtrisées par le cartographe qui souhaiterait exploiter ces données.

1.2.4. Wikidata et Openstreetmap, des sources d'informations volontaires, contributives et libres

Pour analyser les modes de production mis en place afin de créer des informations organisées et logiques, selon des méthodes soutenant la créativité des participants, nous allons prendre l'exemple des projets de contribution Wikidata et OpenStreetMap. Ces deux cadres de production se ressemblent car ils exigent des comportements professionnels et des connaissances expertes de la part de leurs contributeurs, tout en s'efforçant d'accueillir des néophytes dans leur communauté.

Wikidata est un projet de la Wikimedia Foundation (WMF), publié en 2012. Il vise à constituer une base de connaissances librement éditable. Cette base de données est placée sous la licence *Creative Commons Zéro* qui est une reproduction des conditions du dépôt dans le domaine public. L'usage de la base de données n'est restreint par aucune contrainte. L'idée de la construction de la base Wikidata est apparue avec le développement du web sémantique. En effet, Wikidata a eu pour rôle initial d'être la base de connaissances à laquelle sont associés les documents, textuels ou iconographiques, du projet Wikimédia. Cette organisation classifiée de l'information répond à l'objectif du Web sémantique, c'est-à-dire qu'elle vise à rendre accessible, par le moyen des machines, les connaissances humaines. Les connaissances constituées en classes ne sont pas limitées aux seules publications sur Internet. Ainsi, le site de consultation de Wikidata prend-il désormais la forme d'une collection d'articles – parfois autonomes de Wikipédia – relatifs à un individu (ou entité) ou relatifs à une propriété.

Chaque élément de la base Wikidata est référencé par un identifiant unique. Un élément est, soit une entité, soit une propriété. Une entité a un identifiant débutant par la lettre

Q et une propriété a un identifiant débutant par la lettre P. Le formalisme suivi pour renseigner la base de connaissances Wikidata est le modèle RDF pour *Ressource Description Framework*. Développé par le W3C, le modèle RDF a pour rôle de proposer une organisation standardisée et simple d'informations en graphe orienté. Le W3C a proposé ce formalisme en 2004, dans le but de participer à des développements d'outils et de bases de données interopérables.

Le graphe peut être très simple et n'être constitué que de deux nœuds et d'un lien. Il s'agit d'un *triplet* qui décrit une information[75] comme par exemple :

Terre (Q2) (élément) → point culminant (P610) (propriété) → Everest (Q513) (valeur).

Les deux nœuds (Terre et Everest) sont reliés par une propriété. Cette propriété caractérise le premier nœud en ayant pour valeur le second nœud. Ainsi, un fichier RDF est-il un ensemble de triplets. Dans Wikidata, une propriété associée à une entité (point culminant (P610) → Everest (Q513)) forment une déclaration ou *statement*. Grâce aux identifiants uniques mobilisés dans plusieurs triplets, des entités sont décrites par plusieurs déclarations et liées à d'autres entités.

Ce formalisme technique est une réponse au modèle de description par ontologie. Le projet Wikidata est donc un ensemble socio-technique menant à la constitution d'une base de connaissances libre et construite selon le modèle ontologique. Sa formalisation est celle d'un système en graphe. Les connaissances qui y sont enregistrées peuvent soit être singulières (par exemple, Cambridge), soit être le résultat d'une généralisation ou d'une simplification et s'élever au niveau du concept (ville). De ce fait, la définition d'une ontologie commune facilite le partage des connaissances.

La base de données Wikidata peut être techniquement explorée par le biais d'une interface d'interrogation, en usant d'un langage de requêtes, adapté à la formalisation des données stockées. Soulignons que l'accès aux connaissances de la base de données dépend d'abord de l'usage approprié de la syntaxe du langage informatique et ensuite de la compréhension par le lecteur de l'ontologie construite par les contributeurs. Si chacun de ces prérequis est atteint, la base de données est source de nouvelles

75 Comme spécifié dans les recommandations du W3C : https://www.w3.org/2001/sw/wiki/RDF*: This linking structure forms a directed, labeled graph, where the edges represent the named link between two resources, represented by the graph nodes. This graph view is the easiest possible mental model for RDF and is often used in easy-to-understand visual explanations*.

connaissances, dans la mesure où les réponses ne sont pas préparées en amont. Chaque réponse est le résultat d'un extrait du graphe global que constitue la base de connaissances et elle ne dépend donc que de la pertinence de la question posée. Le croisement des classifications, permis par cette organisation de l'information, donne accès à un niveau de savoir davantage comparable aux connaissances tirées de la lecture d'un texte documentaire qu'aux informations tirées d'une base de données relationnelle.

L'utilisation d'une base de connaissances présente des caractéristiques propres à toute base de données. Il est possible de l'éditer et de l'interroger. Ensuite, par le moyen d'une interface d'interrogation avec la base de données, un résultat pourra être proposé au lecteur comme une réponse. La question et la réponse peuvent être simples. Par exemple, on pourra poser la question suivante : quelles sont les instances de la classe « ville » ? La réponse prendra alors la forme d'un graphe. En effet, les entités (les nœuds du graphe) sont reliées par des propriétés (ses arêtes). Il est également possible de transformer ces réponses en un tableau d'individus (les instances) et de caractères (leurs propriétés) dont les valeurs sont les valeurs uniques de chaque propriété. Par exemple, la requête SPARQL ci-contre (fig. 12) a pour résultat un tableau de 83 individus, qui répondent à la classification d'une ville britannique dans Wikidata, comme Cambridge, Birmingham et Liverpool.

```
SELECT ?ville_uk ?ville_ukLabel WHERE {
BIND(wd:Q145 as ?Royaume_Uni).
?ville_uk wdt:P17 ?Royaume_Uni.
?ville_uk wdt:P31 wd:Q515.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
}
```

Figure 12 Requête sur les villes du Royaume-Uni

Cependant, une complexité peut être adjointe à une question simple. On pourra ainsi se demander sans se restreindre à un seul pays quelles instances de la classe ville présentent une population proche à un millier près de la population de Cambridge (fig. 13). Parmi les dix villes qui répondent à la requête citons, Ulm en Allemagne et Léon en Espagne.

```
SELECT ?VilleMondeLabel ?PaysLabel ?PopVille WHERE {
 { SELECT ?Cambridge ?PopCambridge WHERE {
  BIND(wd:Q350 as ?Cambridge).
  ?Cambridge wdt:P1082 ?PopCambridge. } }
 ?VilleMonde wdt:P31 wd:Q515.
 ?VilleMonde wdt:P1082 ?PopVille.
 FILTER (abs(?PopCambridge - ?PopVille) < 1000).
 ?VilleMonde wdt:P17 ?Pays.
  SERVICE wikibase:label { bd:serviceParam wikibase:language "[AUTO_LANGUAGE],en". }
} ORDER BY ?VilleMondeLabel
```

Figure 13 Requête sur la population des villes dans le monde et leur proximité à la ville de Cambridge

La classification est donc le résultat de réflexions sur les concepts et sur les objets spécifiques, associés à ces concepts. Les concepts sont considérés comme les éléments généraux de la classification ; ce sont les classes et les sous-classes de l'ontologie. Chaque élément de l'ontologie, qu'il soit très général ou très particulier, est une entité. Chaque entité doit être identifiable dans son individualité. D'un point de vue formel, chaque entité est reconnue grâce à un identifiant unique dans la base de données. Quant aux instances, il s'agit des entités individuelles d'une classe. Ces instances n'acceptent aucune entité comme une sous-classe. Cependant, bien qu'individuelle, une entité peut partager des propriétés avec d'autres entités qui relient les concepts et les instances (Delestre et Malandain, 2017, p.143).

Cette classification a été mise en œuvre dans la base de données de connaissances Wikidata. Par exemple, le concept de « ville » y est défini par la courte indication ou label : « *large and permanent human settlement by size of its inhabitants* ». Selon les contributeurs à Wikidata, l'entité « ville » est une sous-classe des entités « établissement humain » et « communauté ». Par l'entité « établissement humain », le concept de ville est rattaché au concept plus général « d'objet géographique », placé un rang plus haut dans la classification. De même la sous-classe « communauté » renvoie au concept général de « groupe social ».

La contribution au projet Wikidata est ouverte sans contrainte d'accès. Il est possible de contribuer des propriétés à des entités existantes et de créer des entités en utilisant uniquement son adresse IP comme identifiant de contributeur. Cependant, il semble que les contributeurs s'attachent le plus souvent à s'enregistrer afin de participer, comme

l'indique la lecture des historiques de pages Wikidata, dans lesquels on peut lire le pseudonyme de chaque contributeur à cette page.

Si créer une entité est un geste immédiat et libre, il n'en est pas de même pour les propriétés. Lorsqu'un contributeur exprime le besoin de créer une nouvelle propriété pour un type de connaissance encore non formalisé dans l'ontologie de Wikidata, il doit la proposer à la communauté via le wiki (Wikidata:Property proposal, 2019). Cette proposition doit être accompagnée d'une documentation qui comprend le formalisme et les contraintes de la propriété proposée, des exemples d'applications et la preuve d'une originalité par rapport à l'état de l'ontologie. Comme il est expliqué dans la page du *wiki* présentant les propriétés, après un certain temps (c'est-à-dire après des semaines), si un certain nombre de contributeurs ont approuvé cette proposition et qu'il n'y a pas d'opposition ou qu'elle est très faible, un créateur de propriétés ou un administrateur intégreront la nouvelle propriété au projet[76]. Dans Wikidata, les propriétés présentent des caractéristiques très similaires aux entités, mais elles portent également des contraintes. D'abord, une propriété utilise un type de données, défini à sa création. Par exemple la propriété « date de fondation ou de création » (P571) ne peut être associée qu'à des données de type « point dans le temps »[77].

Comme annoncé précédemment (en début de section), les projets OpenStreetMap et Wikidata présentent des principes initiaux très similaires. Chacun se positionne comme relevant du courant du Libre, construit en communauté de contributeurs aux connaissances et aux compétences variées. Leur visée reste la progression de l'accumulation des connaissances. Les principes de libre accès, de libre édition et de consensus sont des éléments communs à ces deux projets contributifs. Pourtant, les deux projets ont choisi des processus d'itération partiellement divergents, ainsi que des implémentations techniques et des choix méthodologiques radicalement différents. En effet, à l'inverse de Wikidata, OpenStreetMap pose l'amélioration de la folksonomie78 comme libre. S'il n'existe aucune contrainte technique à la création de clés ou de

[76] « *When after some time there are some supporters, but no or very few opponents, the property is created by a property creator or an administrator (see List of property creators).* » cité depuis https://www.wikidata.org/wiki/Help:Properties

[77] Inception, 2019.

[78] Rappelons que le terme folksonomie désigne la pratique de classification d'informations (photographies, pages web, signets) par les utilisateurs d'Internet qui marquent les ressources disponibles sur le réseau par des étiquettes (ou *tags*). La folksonomie est définie par les auteurs, reprenant (Wal, 2005), comme « le résultat de l'activité d'association libre d'étiquettes (*tags*) à des données ou objets (tout ce qui est identifié par une URL) dans le but personnel de les retrouver ultérieurement » (Hombiat et *al.*, 2015, p.106).

valeurs, il est recommandé de les proposer au vote via le wiki et d'accompagner cette proposition d'une documentation. D'ailleurs, ce sont d'abord les clés et les valeurs documentés sur le wiki qui sont mobilisés par les contributeurs (Hall et al., 2018). Cette modeste distinction dans le processus de délibération correspond à un écart méthodologique bien plus conséquent. Les deux modes d'écriture s'opposent. Le projet OpenStreetMap est écrit à l'aide de *tags* tandis que le projet Wikidata recourt à une ontologie formalisée. Hombiat et *al.* (2015) décrivent la folksonomie issue d'OpenStreetMap comme « intrinsèquement plate : il n'existe pas *a priori* de relations hiérarchiques entre les *tags* ». Dans leur article, Hombiat et *al.* expriment leur déception face à ce qu'ils considèrent comme « un obstacle majeur pour le partage et la réutilisation des grands volumes de données » contenus dans la base de données OpenStreetMap. En réponse à ce problème, ils proposent une ontologie des tags d'OSM, afin de parvenir à une exploitation sémantique des données. Ainsi, à l'argument de la simplicité d'une écriture en *tags*, les chercheurs opposent l'idée d'une structure complexe au service de « l'utilisabilité » des données. Pour notre part, nous considérons que la prise en compte du méta-modèle proposé par Hombiat et *al.* repose sur un écart méthodologique et conceptuel trop important pour être intégré par la communauté OpenStreetMap. Le formalisme en ontologie est défendu par ces chercheurs dans le but d'une meilleure utilisabilité[79] des données. Mais les contributeurs à OpenStreetMap ont plutôt comme perspective l'interopérabilité entre la base de données OSM et d'autres bases de connaissances libres. Ainsi, les contributeurs à OpenStreetMap peuvent-ils ajouter le *tag* wikidata=* à tout objet. La page du *wiki* OpenStreetMap consacrée à Wikidata explique l'intérêt de ce tag, en soulignant l'intérêt de l'interopérabilité. L'interopérabilité avec Wikidata permet d'acquérir des informations supplémentaires à celles issues d'OSM, comme des connaissances historiques ou démographiques enregistrées dans Wikidata. La connivence entre contributeurs est visible dans le wiki de Wikidata où on peut lire : « *We particularly welcome OSM contributors who are new to wikidata* ! »[80]. Le partage des principes fondamentaux du Libre, l'interopérabilité des données OSM et Wikidata et la porosité des communautés de contributeurs est également permise par le développement de cadres juridiques communs pour l'utilisation de ces données.

79 L'utilisabilité est analysée dans ce travail de thèse sous l'angle du droit (I.3), de la qualité, de la fiabilité (II) et de la technique (III).

80 Nous souhaitons la bienvenue aux contributeurs OSM qui sont nouveaux dans Wikidata !

1.3. Les droits d'utilisation des données spatiales

L'ouverture des données est un processus de mise en partage des données afin de permettre leur réutilisation. La réutilisation est rendue possible, volontairement, par des cadres juridiques nouveaux. Ces cadres s'appliquent à différents maillages administratifs et se sont imposés au cours des quarante dernières années. En France, en Europe et à l'international, la définition juridique des données ouvertes, est une question récente. Les principaux outils juridiques mis en œuvre sont des lois, des actes réglementaires et des licences (Maurel, 2012). La licence[81] libre est le modèle de protection principal choisi par les défenseurs des productions contributives en ligne. Les premiers modes de protection des ressources publiées pour être partagées en ligne ont été ces licences libres. Une partie d'entre elles est encore aujourd'hui mobilisée afin de protéger le partage de données ouvertes (pas uniquement publiques). Elles ont donc été écrites en dehors de tout contexte administratif national ou supranational. Par la suite, des organisations internationales comme Creative Commons, la Free Software Foundation ou encore l'Open Knowledge International, ont diffusé des licences permettant le partage et la réutilisation de données, tout en assurant une protection juridique. Ces associations ont « façonné en profondeur la dynamique de l'*open data*, notamment en France entre, d'une part, les expérimentations et les politiques d'ouverture des données conduites par les autorités publiques et, d'autre part, les interventions de groupes issus de la société civile » (Maurel, 2012). En outre, ces associations sont soit à l'origine des licences des logiciels libres, soit largement influencées par leurs principes. De ce fait, la liberté défendue par les licences libres consiste en un droit de réutilisation de la création protégée par la licence, par tout autre acteur, sans restriction d'accès. Cette réutilisation est commercialisable. Certaines conditions sont cependant parfois exigées du nouvel utilisateur.

Si les premières licences libres, conçues pour les logiciels, ont pu être utilisées pour protéger des documents, elles n'étaient pas adaptées à la protection de bases de données. En Europe, les documents et les contenus sont protégés par le droit d'auteur. Les bases de données peuvent également être protégées par le droit d'auteur (dans le cas d'une

81 « Les licences libres sont des contrats qui permettent aux auteurs – et à eux seuls – de décider d'accorder des libertés aux utilisateurs des œuvres plutôt que de poser des restrictions. Le créateur peut avec ces instruments choisir une ouverture plus ou moins grande de son œuvre, en maintenant seulement certains des « droits exclusifs » qui s'appliquent normalement en bloc aux œuvres. » (Maurel, 2014)

originalité en termes de modèle de données) mais elles sont davantage protégées « par un droit particulier de propriété intellectuelle, dit *sui generis*, qui permet de s'opposer à certaines actions comme les extractions quantitativement ou qualitativement substantielles de données » (Maurel, 2012). En dehors de l'Union européenne, ce droit dit *sui generis* n'existe pas. Ce sont certaines conditions des droits d'auteur ou du copyright qui sont mobilisées alors pour protéger les données. La licence ODbL, à la manière de plusieurs licences libres désormais, s'appuie sur le droit *sui generis* ainsi que sur le droit d'auteur et vise une efficacité internationale d'application de la licence.

La licence libre est une composante juridique mais également communicationnelle de l'ouverture des données. Par le moyen des licences, les institutions publiques ainsi que les associations de contributeurs affirment une volonté politique de partage des données, sans limite d'accès. En outre, les licences sont une reconnaissance de la valeur des données. En clarifiant les conditions de diffusion de ces données, les licences libres diffusent un discours en parallèle. En mettant en ligne des données publiques selon des formats ouverts et selon des licences autorisant au moins partiellement leur réutilisation, les institutions publiques participent au processus démocratique de critique des administrations et à l'économie numérique. La licence libre est également une solution juridique à l'ouverture des données de la recherche répondant au principe scientifique de reproductibilité des résultats de travaux de recherche.

En ce qui concerne les projets contributifs, la licence libre porte un message à destination des contributeurs. Ainsi, par exemple, les pages d'aide aux premières contributions à Wikipédia, dirigent-elles le lecteur vers les cinq principes fondateurs de Wikipédia. Il y est expliqué que l'ensemble des pages Wikipédia sont publiées sous une licence libre. Cette licence autorise la réutilisation sous certaines conditions. Une de ces conditions exige la mention de Wikipédia, et si possible, des auteurs de la page utilisée. En outre, dans le cadre de cette licence, il est affirmé que « personne n'a le contrôle d'un article en particulier ». Cet aspect de la licence soutient un travail contributif. De manière similaire, en cas de diffusion publique des données, la licence du projet OpenStreetMap exige de mentionner dans les crédits la formule suivante : « © les contributeurs d'OpenStreetMap ».

La licence libre est donc un élément constitutif des projets contributifs en tant que clarification des principes de partage et moyen légal de partage. Dans cette section,

nous allons analyser la diffusion du modèle de protection des œuvres par les licences libres. D'abord conçue à destination de logiciels, la licence libre a été inventée en réaction à des appropriations de codes informatiques, publiés à l'origine sans protection et pour être partagés, améliorés puis repartagés. L'élaboration du concept de licence libre est une réaction de défense face à la « fermeture » de certaines publications. La licence libre est une expression qui sort des cadres classiques juridiques et dont l'invention dans les années 1980 a été décisive car c'est sa définition qui a guidé la constitution du courant du Libre (1.3.1. Construction de licences de distribution libres). Dans les années 1990, les acteurs favorables au partage des codes informatiques, des contenus et des bases de données ont poursuivi le travail de définition des licences. Ils ont utilisé le texte des licences comme des manifestes en faveur du partage pour une production de contenus plus efficace et citoyenne (1.3.2. Acteurs et outils de communication).

1.3.1. Construction de licences de distribution libres

La confidentialité des logiciels, développés entre chercheurs et experts, au début de l'informatique, ne présageait pas de leur future valeur économique jusqu'à la fin des années 1970. Ainsi, les premiers logiciels ont-ils été écrits par des communautés d'experts et diffusés pour être modifiés sans restriction d'accès. La démocratisation de l'informatique a cependant fait évoluer l'intérêt du grand public pour ces moyens technologiques. Au cours des années 1980, les logiciels sont devenus des biens commerciaux attractifs, grâce aux nouvelles fonctionnalités qu'ils offraient déjà et à leur caractère original. Les logiciels sont alors reconnus comme des œuvres pouvant être protégées par le droit d'auteur ou par un copyright.

Aux États-Unis, entre 1960 et 1980, le droit d'auteur sur un logiciel est progressivement défini. Les producteurs de logiciels propriétaires ont développé leurs produits sur cette base juridique, les assurant qu'aucune aliénation de leur création ne serait autorisée. Le système économique de vente de logiciels encapsulés et prêts à l'emploi, dont Microsoft et Oracle sont des acteurs principaux, est toujours d'actualité.

En France, la première étape menant à un cadre législatif propre à l'*open data* est la loi CADA, du 17 juillet 1978, relative à la réutilisation des informations publiques

(Maurel, 2012 ; Clément-Fontaine, 2016). En application de la loi CADA, les administrations publiques ont autorisé la réutilisation d'informations publiques, aux conditions de différentes licences et sans coordination (Clément-Fontaine, 2016). Ensuite, en 2001, la mission Etalab est créée par décret du Premier ministre. Il s'agit d'un nouvel organe de régulation de la publication des informations publiques, qui vient compléter les missions de la Commission d'accès aux documents administratifs (CADA). L'article 2 du décret stipule que « La mission Etalab » est chargée de la création d'un portail unique interministériel, destiné à rassembler et à mettre à disposition, librement, l'ensemble des informations publiques de l'État, de ses établissements publics administratifs et, si elles le souhaitent, des collectivités territoriales et des personnes de droit public ou de droit privé chargées d'une mission de service public »[82]. En octobre 2011, Etalab publie la Licence Ouverte et, en décembre de la même année, Etalab met en ligne le portail data.gouv.fr. Cette nouvelle licence est conçue pour répondre à « une exigence forte de transparence de la donnée et de qualité des sources en rendant obligatoire la mention de la paternité ». La mission Etalab présente sa licence comme « une opportunité de mutualisation pour les autres données publiques en mettant en place un standard réutilisable par les collectivités territoriales, qui souhaiteraient se lancer dans l'ouverture des données publiques »[83]. Lionel Maurel (2012) et Mélanie Clément-Fontaine (2016) précisent que la pratique de l'*open data* a été menée selon une dialectique entre les collectivités publiques, les administrations nationales et des associations. Lionel Maurel signale le rôle critique et constructif des associations et organisations suivantes : Regards Citoyens, Veni Vedi Libri, LiberTIC, Creative Commons France, Wikimédia France.

Au cours des années 1980, une partie des créateurs de logiciels réalisent que leurs créations sont placées sous copyright lorsqu'elles sont intégrées à des publications logicielles, elle-même diffusées par les nouvelles industries du logiciel. Ces créateurs ont alors introduit une nouvelle forme de licence de protection d'une propriété intellectuelle, le *copyleft*. Une licence de type *copyleft* concède le droit d'utiliser, de copier, d'étudier et de modifier l'œuvre licenciée. En contrepartie de ces droits, le logiciel redistribué, modifié ou non, conservera la licence *copyleft*. Le principe de

82 Décret n° 2011-194 du 21 février 2011 portant création d'une mission « Etalab » chargée de la création d'un portail unique interministériel des données publiques.
83 Article « Licence Ouverte / Open Licence » sur le site next.etalab.gouv.fr : http://www.etalab.gouv.fr/etalabpublielalicenceouverteopenlicence. Consulté en mai 2019.

copyleft est conçu en opposition au principe de copyright. Cette condition de protection du bien, conçu comme un bien commun non aliénable, s'intitule « le partage à l'identique » (*share-alike*). Le schéma de production ainsi créé vise à garantir la diffusion de la création sans restriction et à conserver l'objectif d'utilité sociale et intellectuelle d'une diffusion des codes informatiques. Paradoxalement, la faculté de diffuser sous licence libre un logiciel ou toute œuvre de l'esprit, est issue du droit d'auteur ou copyright (selon la législation). Le droit dont dispose l'auteur de choisir les modalités de réutilisation de sa création lui permet de décider des conditions selon lesquelles l'œuvre peut être diffusée (Xifaras, 2010). Un auteur peut mobiliser ses droits d'auteur pour les céder à d'autres auteurs.

L'invention du *copyleft* et l'usage des licences dans le contexte du Libre sont des procédés de défense face au risque d'aliénation des créations libres. Le type de licence *copyleft* est une solution juridique, qui permet à un auteur de céder ses droits à d'autres utilisateurs et d'interdire tout usage restrictif par réappropriation. La Licence publique générale GNU (GNU General Public License, GNU GPL), proposée en 1989 par Richard Stallman pour distribuer les programmes créés dans le cadre du projet GNU, est la première licence *copyleft* à largement se diffuser. Cette diffusion est de l'ordre de l'usage de la technologie du projet GNU, mais recouvre aussi la diffusion d'un mode de pensée. Le manifeste GNU, publié en 1985 par Richard Stallman, soutenu par d'autres développeurs du projet GNU, expose le principe suivant : « GNU n'est pas dans le domaine public. Tout le monde aura le droit de modifier et redistribuer GNU, mais aucun distributeur ne pourra restreindre ces futures redistributions. C'est-à-dire que les modifications propriétaires seront interdites. Je veux m'assurer que toutes les versions de GNU resteront libres »[84]. Ce texte reste la référence théorique principale du mouvement (Xifaras, 2010).

Le concept de *copyleft* est donc apparu en réaction à l'appropriation de certaines réalisations dont la propriété n'était pas déclarée par leurs auteurs. Le concept, plus tardif, de licences libres désigne une catégorie plus large de licences (et intègre la forme *copyleft*) car il repose uniquement sur les quatre libertés que sont le droit d'utiliser, de

84 Traduction de « *GNU is not in the public domain. Everyone will be permitted to modify and redistribute GNU, but no distributor will be allowed to restrict its further redistribution. That is to say, proprietary modifications will not be allowed. I want to make sure that all versions of GNU remain free.* » Consulté en ligne en avril 2019 : https://www.gnu.org/gnu/manifesto.fr.html

copier, d'étudier et de modifier l'œuvre[85]. Une licence libre n'impose *a priori* aucune restriction à l'exploitation de l'œuvre d'esprit. Les licences libres sans restriction peuvent permettre, lorsqu'aucune restriction n'y est adjointe, une totale liberté, s'approchant du domaine public. Ces licences libres sans restriction sont qualifiées de permissives, montrant ainsi une forme d'abandon par le créateur de tout contrôle sur la diffusion de la création. Ainsi, à l'inverse d'une *copyleft*, qui vise à garantir les quatre libertés, quelle que soit l'évolution de l'œuvre, une licence permissive ne garantit pas légalement la continuité d'une diffusion accessible à tous. Cependant, l'attribution de l'auteur de l'œuvre est le plus souvent exigée par les licences permissives. De même, il existe différentes formes de licences *copyleft*. Une distinction entre *copyleft* standard et *copyleft* fort est parfois proposée. Dans le cas d'un *copyleft* standard, l'œuvre et ses modifications directes ne peuvent être diffusées sous une autre licence que celle définie au départ. Dans le cas d'un *copyleft* fort, les œuvres dérivées de l'œuvre initiale devront également être diffusées sous la même licence.

Les premières licences libres ont donc été d'abord conçues afin de protéger des logiciels, des codes informatiques et leurs documentations. Avec le développement d'Internet, tous les types de contenus comme les photographies, les vidéos, les textes ainsi que les bases de données (en tant qu'architecture d'information originale et en tant que collections de contenus d'informations) ont progressivement pu être protégés par des licences libres. Le partage de données sous licence libre a commencé à la fin des années 1990 avec des projets comme l'Open Directory Project et Tela Botanica. Par la suite, c'est le projet OpenStreetMap qui est devenu, dans le domaine des bases de données, un modèle de partage et de protection légale. D'abord sous licence Creative Commons By SA, le projet assure désormais une protection spécifique, adaptée à la forme de la base de données, grâce à l'Open Database Licence (ODbL). Cette licence impose que toute base de données, issue de la base de données OSM, modifiée ou non, soit diffusée sous la licence ODbL. Cette licence est de type *copyleft* et assure à ses contributeurs la continuité d'une diffusion libre des informations créées.

Le 12 septembre 2012, le projet OSM, au travers de l'OpenStreetMap Foundation, a substitué la licence Open Database (ODbL) à la licence CC BY-SA, alors en vigueur pour sa base de données. Ce changement de licence a été le résultat d'un processus

85 Les quatre principes du Libre par la Free Software Foundation : « « *What is Free Software? - GNU Project - Free Software Foundation (FSF)* » . Consulté en mai 2019.

commencé courant 2009, au cours duquel les contributeurs actifs dans la gouvernance du projet OSM ont débattu démocratiquement de l'intérêt de l'ODbL. Le 5 décembre 2009, après dix mois de travail entre le groupe de travail sur les licences de la Fondation OSM et les experts d'Open Data Commons, le conseil (*Board*) a proposé au vote l'application d'un plan de transition pour passer de la licence de la base de données OSM à la licence ODbL 1.0. La motivation principale d'un tel changement était une inadéquation de la licence CC BY-SA 2.0 avec les besoins du projet OpenStreetMap. En effet, la Licence CC-BY-SA 2.0 protégeait les œuvres artistiques telles que des livres et des photographies. Cependant, son utilisation pour la base de données OSM a entrainé des débats d'interprétation. Si une carte présentant des données OSM était intégrée dans un livre, l'ouvrage était-il alors sous licence CC par la clause *Share-Alike* ? Le risque de se placer dans l'illégalité du fait d'une lecture plus ou moins littérale des conditions d'utilisation a été atténué par l'emploi de la licence ODbL. Si les données dérivées d'OSM doivent être placées sous la licence ODbL, d'autres licences ouvertes ou privatives peuvent être choisies pour les œuvres produites à partir de la base de données (*Produced Work*). Par exemple, les tuiles d'un service de cartographie en ligne peuvent être protégées par une licence autre que la licence ODbL. Toute donnée de la base principale d'OSM, créée avant cette date, restait sous l'ancienne licence, le temps d'obtenir l'accord du contributeur pour le changement de licence. Les données issues de contributeurs qui n'avaient pas donné leur accord ont été retirées de la base publique et archivées sous le régime de CC-BY-SA 2.0. Leurs producteurs peuvent cependant accepter la licence à tout moment.

La licence ODbL[86] protège exclusivement les données et les bases de données. Elle conserve dans son principe la condition *Share-Alike* de la licence Creative Commons, très protectrice mais cette condition s'applique seulement aux biens que sont les bases de données. L'ODbL est de type *copyleft*. En effet, deux concepts d'utilisation des données sont bien distincts : la « base de données dérivée »[87] et la « Création Produite »[88]. La base de données dérivée consiste en une base de données modifiée ou adaptée depuis la base de données initiale OpenStreetMap, dont les transformations concernent une partie substantielle du contenu de la base de données. Pour toute base

86 Dont la documentation est accessible à cette adresse : http://opendatacommons.org/licenses/odbl/
87 Traduction de *Derivative database* par Veni, Vidi, Libri. Consulté en ligne en avril 2019 : https://vvlibri.org/fr/licence/odbl-10/legalcode/unofficial
88 Traduction *de Produced work* par Veni, Vidi, Libri. Consulté en ligne en avril 2019 : https://vvlibri.org/fr/licence/odbl-10/legalcode/unofficial

de données dérivées, la clause de « partage à l'identique », selon la licence ODbL dans ce cas, s'applique. Au contraire, la « création produite » peut-être diffusée sous toute autre licence car il s'agit d'une création produite via un support visuel, audiovisuel ou audio, ou encore littéraire. Lors d'une création, si les données de la base initiale ont été modifiées, les données doivent être diffusées « selon la licence ODbL » et la mention de la source des données doit être intégrée à la création. La licence ODbL implique l'attribution de la paternité des données, leur repartage aux conditions identiques après modification en cas d'utilisation publique, sans aucune restriction technique à leur accès. Ces conditions sont plus connues sous leur appellation anglaise, dans le résumé de la licence, diffusé par Open Data Commons. Elles sont intitulées respectivement : *Attribute, Share-Alike* et *Keep Open*. Chacun est libre de partager (*To Share*), c'est-à-dire de copier et de distribuer les données, de créer une autre production (*To Create*) à partir de la base de données et de l'adapter (*To Adapt*) par modification ou transformation. Si une personne choisit de rendre publique une version non modifiée des données OSM, elle est tenue d'attribuer les données aux « Contributeurs d'OpenStreetMap ». Si une personne rend publique une version modifiée des données OSM, elle doit partager sa production sous la licence ODbL, sans en restreindre l'accès. L'utilisateur peut commercialiser les données. Cependant, celles-ci étant dérivées d'OSM, elles sont sous licence ODbL. Toute autre personne peut donc également commercialiser les mêmes données ou les redistribuer gratuitement.

On notera que le respect des licences est également attendu de la part des contributeurs. Les sources externes telles que Google Maps ou TomTom ne peuvent être utilisées comme références car leurs propres licences l'interdisent. Dans le projet OSM, la source la plus encouragée reste celle de l'expérience du terrain. Cependant, plusieurs sources des différentes licences peuvent être compatibles et utiles au projet. Un certain nombre des données disponibles dans la base de données OSM ont été reversées par des collectivités locales ou des organismes favorables à l'ouverture des données. Ces données versées dans la base relèvent du domaine public ou de licences compatibles avec la licence ODbL. Il faut noter que la licence choisie par OpenStreetMap n'est pas compatible avec d'autres licences ouvertes, dont les conditions moins exigeantes.

L'acteur principal de ce changement de licence est l'OpenStreetMap Foundation. Comme cela est expliqué sur la page du *wiki* d'Openstreetmap, la Fondation OpenStreetMap est une association britannique dont le siège est à Cambridge sans but

lucratif. Créée en 2006, la Fondation est responsable des serveurs et est propriétaire du nom de domaine www.openstreetmap.org. La Fondation a pour rôle d'être la personne morale collectant les fonds de soutien au projet. Elle joue également un rôle de représentation auprès d'institutions et d'acteurs privés. Enfin, elle est chargée de garantir l'application de la licence ODbL[89]. La Fondation défend l'idée qu'une licence *copyleft* assure la croissance et la préservation de la base de données OpenStreetMap, conçue comme un bien commun de la connaissance. La clause « partage à l'identique » sert cette intention d'accroissement du bien commun. Par ailleurs, la Fondation vise à encourager l'usage de la base de données. Les usages de la base de données OSM impliquent, le plus souvent, des modifications. Afin de soutenir l'usage de la base de données, sans laisser de doute aux conditions de repartage de ces modifications, la Fondation a proposé des lignes directrices dites *Community Guidelines*[90] permettant de préciser les règles d'application menant à une base de données dérivée (*derivative database*). Celles-ci visent à distinguer des modifications dites triviales des modifications ayant pour but d'améliorer la base de données. Les modifications triviales n'ont pas à être diffusées sous la licence ODbL. Sur la page des *Community Guidelines* de la Fondation, il est expliqué que celles-ci sont nécessaires, compte tenu du caractère récent des lois sur lesquelles la licence ODbL est fondée. Ainsi, est triviale une modification qui n'est pas utile à l'amélioration de la base de données source[91]. Plusieurs exemples sont proposés dans cet article. On retiendra que la re-projection des données, la transformation en graphe pour le calcul d'itinéraire ou la compression des données sont des transformations triviales, n'impliquant pas l'application de la clause « partage à l'identique ».

89 « *The OpenStreetMap Foundation is there to protect the OSM data to keep it Free and Open*. » Une des missions principales de la Fondation, déclarée sur son site : https://wiki.osmfoundation.org/wiki/Mission_Statement : consultée en novembre 2018.

90 Celles-ci s'appliquent à l'usage de la base de données OpenStreetMap et non sur toute base de données publiée sous la licence ODbL.

91 « *We therefore define a term "trivial transformation", (for want of a better phrase!) which covers alterations of OpenStreetMap data which are not considered interesting or useful enough to warrant the conditions of a derivative database. The word « trivial« should not be interpreted as technically trivial, but is intended to be "trivial" in the sense of modifications or additions to the data. For example, adding or correcting* data would not ever be considered "trivial » (OSM foundation , Licence/Community Guidelines/Trivial Transformations – Guideline, Consulté en mai 2019)

1.3.2. Acteurs et outils de communication

La définition des principes fondamentaux des licences libres pour les œuvres logicielles s'est étendue sur vingt ans. La période entre 1960 et 1980 a été le temps de la démocratisation des outils informatiques dans les milieux professionnels experts. Dès lors, une partie de la population élargie au-delà des seuls concepteurs informatiques s'est intéressée à la technicité du développement informatique d'une part, et aux valeurs d'éthique, de protection de la vie privée et de création en commun qu'incarnent le courant du Libre, d'autre part.

Afin de soutenir un développement libre d'un logiciel, il convient d'améliorer sa documentation en même temps que le code du logiciel. De ce fait, la documentation associée au logiciel devient un élément clé pour le bon fonctionnement d'une communauté créatrice d'un code informatique. Jusqu'à la fin des années 1990, la licence du logiciel était étendue à sa documentation. Au début des années 2000, la GNU Free Documentation License (GNU FDL) est la licence de la documentation du projet de système d'exploitation GNU, soutenu par la Free Software Foundation. D'autres licences ont succédé à la GNU FDL. Grâce à ces licences de documentation, les concepteurs visent à garantir la contribution des utilisateurs et des développeurs à l'écriture de la documentation. Les textes de licences et les documentations qui accompagnent les créations deviennent alors des moyens de communication. Par ailleurs, l'accroissement et la clarification des documentations contribuent à la diffusion du mode de production d'œuvres libres et à l'explication des motivations qui guident cette production.

En 2001, a été fondée l'association à but non lucratif Creative Commons, dont le siège est en Californie. L'association a pour but de proposer des contrats-types pour la protection des modalités de partage des œuvres. La même année, le projet d'une encyclopédie libre et contributive en ligne, Wikipédia est publié en ligne sous la licence GNU GPL. En 2003, la fondation Wikimedia, dont le siège est aujourd'hui en Californie, est à son tour fondée afin d'administrer les projets comme Wikipédia, Wikitionnaire ou Wikidata. En 2003, est fondée également la Fondation Mozilla, organisation à but non lucratif, porteuse de l'idée qu'Internet peut être pratiqué dans le respect de la vie privée. La même année, la directive 2003/98/CE du Parlement européen et du Conseil sur la réutilisation des informations du secteur public fixe des

règles facilitatrices pour l'accès des personnes aux documents administratifs. Ubuntu, système d'exploitation basé sur des technologies aux licences libres, est publié en 2004 et devient, au cours des quinze dernières années, un des systèmes d'exploitation les plus employés par les particuliers, par les acteurs du privé et par les institutions publiques. En 2004 encore, le projet Openstreetmap, installé à l'Université de Londres, est mis en ligne et sa base de données est protégée par la licence Creative Commons BY SA (attribution et partage à l'identique). Cette même année, *l'Open Knowledge International* (alors l'*Open Knowledge Foundation*) est fondée. Cette organisation, concentrée sur la valorisation de la production des données ouvertes, a accompagné la Fondation OpenStreetMap dans son changement de licence. Depuis 2011, la France participe, au travers de la mission Etalab, à l'ouverture des données publiques, en proposant la Licence Ouverte dont la condition unique est **l'attribution**. Depuis 2012, l'accès à la base de données OpenStreetMap est protégé par l'*Open Database Licence* (ODbL), conçue par *l'Open Data Commons*, partie de *l'Open Knowledge Foundation*. Chacune de ces dates est un jalon vers des habitudes de partage d'informations plus systématiques.

Au cours de ces quinze dernières années, ces organismes, dont l'objectif est de populariser le courant du Libre, développent des procédés de communication fondés sur une iconographie simple et de brefs textes expressément confectionnés pour être diffusés en ligne. C'est ce rôle que jouent les logos de la Creative Commons. L'association vise à rendre accessible à tous un langage juridique, afin d'encourager le choix volontaire de licences. On peut lire à la page Licence du site de l'association que « puisque la plupart des créateurs, des éducateurs et des chercheurs ne sont pas juristes, nous proposons également les licences dans un format accessible à tous »[92]. Dans les années 2010, le principe du partage de connaissances au service d'une plus grande production de savoirs s'est largement diffusé mais le nombre de licences s'est également fortement accru. L'objectif des organisations les plus actives dans la définition des clauses des licences libres est désormais de lutter contre la « prolifération des licences »[93]. Ainsi, le site Web choosealicence.com (fig. 14) s'ouvre-t-il sur trois

92 https://creativecommons.org/licenses/?lang=fr-FR, consulté en mai 2019

93 On trouve cette expression dans les textes d'Etalab, par exemple. La prolifération des licences a permis l'amélioration de certaines clauses ou leur apparition pour répondre à de nouveaux besoins. Cependant, les organismes comme Etalab, Creative Commons, la Free Software Foundation, ou l'Open Data Commons défendent l'usage d'un nombre limité de licences dont l'interopérabilité ou l'absence d'interopérabilité seraient clairement lisibles pour tout utilisateur.

catégories de besoins. Dans le premier cas, est valorisé le travail itératif en communauté autour d'un code source. Dans le deuxième cas, c'est la simplicité reconnue de la licence MIT qui est mise en valeur tandis que le troisième cas est l'expression de la clause « partage à l'identique » du *copyleft*. Les options esthétiquement présentées sur cette page d'accueil ne sont pas limitées aux logiciels. Un hyperlien redirige, au travers du titre « mon projet n'est pas un logiciel » vers une liste de licences.

Figure 14 Accompagner le choix d'une licence par des textes simples et une conception graphique organisée

Les licences diffusées par la Free Software Foundation, Creative Commons et certaines institutions publiques comme le MIT et Etalab font référence. Deux modalités confèrent à ces acteurs leur légitimité : leurs licences incarnent le courant du Libre et elles répondent aux besoins des producteurs. Le premier élément de crédibilité conféré à une licence libre est sa reconnaissance par des acteurs de référence. Trois références participent de cette reconnaissance : la liste des Licences de logiciel libre compatibles avec la GPL diffusée par la Free Software Foundation, l'Open Source Definition de

l'Open Source Initiative et l'Open Definition de l'Open Knowledge Foundation Network (OKFN). Ces trois références posent des principes à destination des utilisateurs de licences. Les trois organismes partagent des bases communes mais ne reconnaissent pas un même corpus de licence. Si pour l'OKFN, l'Open Definition est davantage un outil au service de la diffusion des concepts d'*open data*, les définitions proposées par la FSF (et sa licence GNU GPL) et l'OSI sont des références pour le choix d'une licence. Par exemple, la majorité des licences référencées dans Wikipédia sont rapportées à ces deux évaluations (voir fig. 15).

Figure 15 Évaluation de la compatibilité de la licence Apache sur Wikipédia

La FSF et l'OSI ont donc un rôle de gouvernance assurant une autorité dans les domaines respectifs du Libre et de l'*open source*. Ce rôle de référence et de contrôle du principe du Libre est également celui d'Etalab pour les administrations publiques françaises. Il ne s'agit pas uniquement d'une garantie éthique mais également d'une garantie fonctionnelle. Comme on peut le lire dans la page relative aux licences du site Web d'Etalab, les licences de diffusions de données et de codes informatiques « prolifèrent »[94]. Ainsi, l'État français a-t-il prévu, au travers de la loi pour une République numérique (2016), la création d'une liste, fixée par décret, de licences qui peuvent être utilisées par les administrations pour la réutilisation à titre gratuit de leurs informations publiques. En outre, cette volonté de restreindre le nombre de licences à quelques licences de référence, vise à faciliter leur compatibilité mutuelle. Un paragraphe de la Licence Ouverte version 2 précise : « La présente licence a été conçue pour être compatible avec toute licence libre qui exige au moins la mention de paternité

94 Article « Licences de réutilisation » sur le site data.gouv.fr : https://www.data.gouv.fr/fr/licences, consulté en mai 2019.

et notamment avec la version antérieure de la présente licence ainsi qu'avec les licences « Open Government Licence » (OGL) du Royaume-Uni, « Creative Commons Attribution » (CC-BY) de Creative Commons et « Open Data Commons Attribution » (ODC-BY) de l'Open Knowledge Foundation ». La licence ODbL a également été adoptée comme une des licences pouvant être applicables aux informations publiques comme « Licence avec obligation de partage à l'identique ». La Licence Ouverte est une licence permissive recommandée pour la majorité des informations publiques. La licence ODbL peut, quant à elle, être employée « si les données publiées répondent à un objectif d'intérêt général »[95]. Cet objectif mis en valeur par Etalab correspond à l'objectif initial du *copyleft*, visant sans qu'Etalab l'ait formalisé ici, à assurer la protection des communs de la connaissance.

La légitimité des acteurs et par extension de leurs licences ne tient pourtant pas uniquement à son degré de compatibilité. En effet, c'est l'efficacité de la licence qui prévaut. Son efficacité repose sur la définition de dispositifs juridiques qui doivent être adaptés à l'œuvre licenciée et qui sont également valables pour le défendeur, en cas de tentative d'appropriation. Le non-respect de différents paramètres peut mener à une violation de licence. Le paramètre le plus courant est celui de la mention de l'origine des données. En cas de non-respect, on parle d'absence d'attribution. C'est la modalité la moins respectée. Les circonstances de la violation prennent des ampleurs différentes en fonction de l'acteur qui publie les données non attribuées. Dans le cas d'OpenStreetMap, une page du *wiki* OSM intitulée « *Lacking proper attribution* » recense les cas identifiés d'absence d'attribution. On reconnaît dans cette page, en introduction, la méthode principale de gestion d'une situation de violation de la licence. Il est recommandé à toute personne identifiant une violation de contacter le diffuseur par les moyens de contacts directs comme les mails. En effet, comme il est indiqué en introduction, le non-respect de la licence tient le plus souvent à une méconnaissance des règles. On remarque d'ailleurs dans la liste de nombreux cas d'utilisation des données OSM avec le service cartographique Leaflet, pour lesquels l'attribution « © les contributeurs d'OpenStreetMap » est manquante. Dans cette configuration erronée, il pourrait sembler à un utilisateur que Leaflet est la source des données. En effet, le service Leaflet offre un moyen simple d'ajouter une carte dynamique à une page Web ;

95Article « Licences de réutilisation » sur le site data.gouv.fr. https://www.etalab.gouv.fr/comment-etalab-aide-les-administrations-a-ouvrir-leurs-donnees, consulté en mai 2019.

son ensemble d'outils intègre alors automatiquement l'attribution à Leaflet et non à OpenStreetMap. En cas d'absence de réponse après une prise de contact directe, il est recommandé d'ajouter la situation de violation de la licence à la liste de la page du wiki. Cette action est un premier signalement du diffuseur, présenté comme non respectueux des conditions d'utilisation. Les procédures judiciaires pouvant être coûteuses et pouvant nuire à la réputation d'OpenStreetMap en tant que diffuseur de données libres, c'est cette méthode de dénonciation qui est préférée. Dans un troisième temps, lorsque la violation de la licence semble volontaire et concerner une quantité importante de la base de données OpenStreetMap, la Fondation devient l'interlocuteur auprès du diffuseur auquel on adresse une critique. Les échanges sont non publics mais une partie d'entre eux est diffusée sur les réseaux sociaux. Par exemple, les figures 16 et 17 correspondent à la publication d'un tweet de mai 2012, célébrant le succès de la Fondation OSM, qui a exigé et obtenu la mention OSM comme source dans iPhoto[96].

Figure 16: Le « succès » de la Fondation OpenStreetMap pour obtenir la mention de la source

96 https://twitter.com/openstreetmap/status/198101512201834497

Figure 17 le fichier intitulé "success.jpg" cité dans le tweet ci-contre

Ainsi, la légitimité des licences repose-t-elle sur trois éléments : leur clarté pour la compréhension par tout utilisateur, l'applicabilité de leurs conditions et la possibilité pour un utilisateur, n'ayant pas formellement respecté la licence, de se corriger. Ce dernier point est représentatif d'une pratique générale d'intégration dans une communauté. Comme il est signalé dans l'introduction de la page du wiki OpenStreetMap : « *Lacking proper attribution* », le but du projet OSM est de produire des données libres et non de débattre sur des problèmes légaux[97]. Cependant, les deux premiers éléments que sont la clarté du texte juridique et l'applicabilité des conditions sont, quant à elles, de plus en plus fréquemment assurées par des professionnels du droit. Par exemple, le Software Freedom Law Center est une association de conseils juridiques soutenant les projets libres. Cette association a participé à la définition de la troisième version de la Licence Publique Générale. De la même manière, l'aura acquise par l'association Creative Commons repose sur le développement, dans plusieurs pays,

97 La page Lacking proper attribution (2019) du *wiki* OSM précise : « *Be gentle! We are a project for providing free maps and not to argue with somebody about legal issues. It doesn't help anybody playing the police and collecting as much cases as possible. This could give OSM a bad image. Don't be too correct. If they give it their best then it is ok.* »

d'un réseau d'experts en droit. L'usage large de ces licences par de grands acteurs d'Internet apporte de la crédibilité à l'association Creative Commons.

Les licences libres sont des contrats par lesquels les auteurs accordent des libertés. Des restrictions peuvent être ajoutées, faisant de la licence libre une licence libre de type *copyleft*, c'est-à-dire assurant un repartage de l'œuvre si elle devait être modifiée et publiée. Par leur originalité, celle d'offrir des libertés plutôt que d'imposer des conditions d'utilisation, les licences libres sont des vecteurs de communication d'une organisation autour de biens communs des connaissances. Les licences libres sont composées de documents juridiques, mais elles sont également accompagnées de documents pédagogiques et de communication. À partir des années 2010, les logos de la Creative Commons sont largement connus ; la licence MIT est choisie pour sa crédibilité dans la pratique du développement informatique[98] et l'appellation *open data* trouve une place dans les lois françaises et dans les directives européennes. Par commodité et pour se différencier les unes des autres, les licences ont pris le nom des projets pour lesquels elles ont été créées. Ainsi, les licences développées dans le cadre du projet GNU sont-elles mobilisées et affichées par des projets comme Linux, Firefox ou Qgis. Les communautés et les acteurs majeurs des contenus libres ont détourné un document, relevant à l'origine uniquement de l'ordre de l'expertise juridique, afin de produire un support de communication en faveur des communs de la connaissance. Alors que la licence libre initiale était un moyen de défense face à l'aliénation des codes informatiques, la licence libre actuelle est un instrument de crédibilité présentant des projets libres comme des alternatives à des créations fermées ou dites propriétaires. Aujourd'hui, cette crédibilité est réduite du fait des incompatibilités entre les licences mais surtout du fait de leur diversité. La licence reste un objet complexe à saisir. Si l'objet juridique ne doit pas être simplifié, car les lois exigent que le texte de licence prévoie des conditions d'applications diverses, un nombre plus réduit de licences apporterait une meilleure interopérabilité entre les données.

98 Comme l'outil node.js, la bibliothèque GDAL ou encore l'émulateur de commande Putty.

Conclusion du chapitre 1

L'information géographique est, par définition, modélisée suite à une réflexion systémique. Une information géographique n'est pas conçue unitairement. Elle est produite conjointement à d'autres informations qui sont interconnectées dans leurs concepts. Dans notre contexte concentré dans la sphère du partage de données ouvertes, l'information géographique est conçue pour être enregistrée dans une base de données géographique. Ce caractère géographique et non seulement spatial d'une base de données tient au fait que les informations qu'elle contient ont été conçues comme faisant partie d'un système. Les informations géographiques sont **inter-reliées** par leurs rapports topologiques mais également par une réflexion plus générale fondée sur les concepts de la géographie et des sciences de l'information. En effet, en intégrant l'écriture d'une base de données géographique à un cadre théorique plus large, la base répondra d'une part à des besoins géographiques en matière de données hydrologiques, démographiques ou de transports, par exemple, et servira d'autre part l'objectif d'interopérabilité (Joliveau, 2011) entre les différentes bases.

L'interopérabilité est un défi technique (compatibilité des plateformes, des langages d'enregistrements des données, des versions de systèmes) et méthodologique (concordance ontologique, gestion des mises à jour, pratique de l'ouverture des données) qui se révèle transverse dans la production de données. L'objectif ne sera pas atteint dans les prochaines années car il nécessite une rationalisation des bases de données déjà existantes. L'interopérabilité servira l'un des objectifs premiers de la production de l'information, c'est-à-dire l'économie de temps et de coût de production et d'exploitation. Par ailleurs, l'objectif prioritaire, d'une organisation de l'information en systèmes, et déjà atteint de la production d'informations, est la déduction de nouvelles informations à partir d'informations existantes. Enfin, l'interopérabilité garantira un usage plus étendu des informations stockées. L'interopérabilité potentielle dans l'univers du Libre et de *l'open data* s'étend avec l'accroissement des projets ouverts. Ainsi comme nous l'avons vu, les contributeurs à OSM intègrent des clés dans la base de données permettant de lier des objets de la base aux objets équivalents des bases Wikipédia (`wikipedia=wikipedia key`) et Wikidata (`wikidata=wikidata key`). **Le processus de production d'informations géographiques repose donc sur un ensemble de choix et sur un cadre**

méthodologique assurant une abstraction plus ou moins forte. Ces étapes doivent être formalisées pour que l'information soit diffusée.

Le défi de l'interopérabilité est le résultat du changement de paradigme qu'a connu la cartographie à partir des années 1960. D'abord, on a pratiqué une cartographie informatisée, sans réellement tirer parti des potentialités de l'informatisation. Puis, au terme du développement des nouvelles normes de partage d'informations et de production en groupes *via* Internet, le numérique impose sa logique. Des technologies adaptées à un web toujours plus interactif ont offert un contexte de travail en groupe à une échelle encore jamais atteinte. Par ailleurs, ces groupes de contributeurs ont progressivement été composés de personnes formées dans les universités et les instituts de technologies. Ces experts des nouveaux enjeux de productions ont, en partie, bâti les principes du Libre. Les experts et les contributeurs motivés par ces principes se sont alors constitués en communautés. **Ces communautés, à la base des systèmes libres, ont établi des règles de partage des connaissances mais également des règles de protection de ces connaissances.** L'étendue du Web a imposé à ces groupes de construire une gouvernance[99] originale et adaptée pour garantir le caractère *commun* de la ressource dans un Internet démocratique. OpenStreetMap est par exemple le résultat des deux bouleversements qu'a connu la production d'informations géographiques et par lesquels elle a commencé à devenir interopérable et ouverte. Dans ce contexte, pour y répondre, « l'utilisabilité » des connaissances est devenu un objectif fondamental. Le mode d'édition en *wikis* a été choisi afin de documenter les projets mais également de favoriser le débat. Des organisations ont été fondées afin de proposer des normes d'écriture de données et de leur documentation. Ces normes concernent à la fois l'implémentation technique des bases de données et les langages informatiques. Accompagnant ces normes, des règles de formalisme se sont imposées et l'exigence de documentations claires a été reconnu comme un paramètre favorable aux projets contributifs fonctionnels.

Dans ce chapitre, nous avons décrit les modalités désormais traditionnelles de production d'informations géographiques. Des années 1980 aux années 2010, des protocoles d'écriture et de diffusion de l'information produite ont été formalisés par différents organes comme la recherche (Sandrine Balley, Colette Cauvin, Howard

99 Au sens d'Elinor Ostrom c'est-à-dire : « un système de règles et normes, définissant les droits (et obligations) attribués aux différents membres d'une communauté » (Oliver Weinstein dans Orsi et *al.*, 2017).

Fisher, Michael Goodchild, Thierry Joliveau, David Maguire), les instituts nationaux géographiques (IGN, USGS, ou la NASA) l'Organisation internationale de normalisation (ISO) et des entreprises (Microsoft, Oracle, HERE, ESRI ou TomTom).

Depuis la fin des années 1990, le nouveau mode de production contributif et libre s'est affirmé. À l'inverse du mode de production institutionnel, le nouveau mode de production libre est autonome de toute responsabilité d'intérêt public. Ce processus est également dénué de tout objectif marchand, contrairement au mode de production des entreprises commercialisant des bases de données géographiques. La production libre présente donc des motivations de construction et de diffusion de l'information différentes du modèle traditionnel. Les principes qui ont mené à la fondation de ces projets sont le libre partage et la libre édition au service de la production de connaissances. Les autorités actuelles préfèrent une production cloisonnée et une diffusion contrôlée aux principes du Libre. Les projets contributifs libres ont donc établi des règles au service du travail collectif dans le but de produire, en groupe, des informations utilisables par d'autres. C'est pourquoi nous avons étudié, dans ce chapitre, les facteurs explicatifs de la constitution des projets contributifs et les modalités de préservation des ressources que sont les licences *copyleft*.

L'infrastructure de la base de données OSM, majoritairement éditée par des contributeurs experts, respecte des protocoles techniques de partage d'informations géographiques désormais traditionnels. Les données sont manipulables depuis tout logiciel de SIG. La documentation est diffusée par le moyen d'un wiki, socle du projet de la base de données OpenStreetMap. Il en découle que les contributeurs s'approprient partiellement le modèle traditionnel par le respect des normes (protocoles de transferts des données) et la documentation même s'ils s'affranchissent d'autres exigences moins tributaires d'Internet. De fait, le modèle traditionnel exige la couverture homogène d'un territoire, la collecte et l'enregistrement des informations selon des règles strictes de cohérence et l'engagement par le producteur d'une vérification systématique de ses publications. En s'appropriant les principes du Libre, le projet OSM se positionne à l'inverse de chacune de ces attentes. **Non seulement le projet OSM ne respecte pas les règles habituelles, mais il en a créé de nouvelles dans le but de rendre la base de données éditable par tous.** Inconvénient : elle est de ce fait plus vulnérable car exposée à des actes de malveillance.

Les contributeurs d'OSM se sont montrés inventifs en mobilisant Internet à leur manière, afin de s'affranchir des autorités institutionnelles et de publier leurs propres productions. La base de données OSM est aujourd'hui une des principales sources de données géographique, notamment dans la production de cartes numériques. Pourtant, du fait du déficit de confiance des utilisateurs son usage est trop souvent réduit à celui du dessin d'un fond de carte pour la localisation de points d'intérêts. Il est vrai que sans contraintes de résultats, les projets contributifs libres ne semblent pas avoir établi de procédure garantissant la fiabilité de leur production. Le mode de production de la base de données OpenStreetMap n'inspire pas la même confiance que certaines institutions auxquelles on délègue l'examen critique de leur production. À partir de l'état de crédibilité actuelle qu'il est possible d'accorder à la base OSM ainsi que l'évolution envisageable de la production de connaissances au sein d'OSM, nous proposerons une prévision de l'évolution de cette crédibilité. **La confiance en la base de données OSM est l'objet du chapitre 2.**

Chapitre 2 : Fiabilité de la base de données OpenStreetMap

Traditionnellement, la fiabilité des publications dites de référence de travaux est garantie par un ensemble de mécanismes culturels ou institutionnels. L'évaluation par les pairs caractérise la science moderne. La relecture et la correction manuelle sont en principe assurées par les diffuseurs d'informations, qu'ils soient producteurs de bases de données ou éditeurs de livres ou de revues (Stvilia et *al.*, 2007). Si cette garantie de fiabilité n'est pas remise en question en tant que telle, ces mécanismes ne correspondent pas aux productions conçues et réalisées dans le cadre du Libre et d'OSM.

Les données de la base de données OpenStreetMap sont librement éditables par tous, à tout moment. Elles sont également librement utilisables sans restriction d'accès. Ainsi, une édition des données par un contributeur *anonyme* est immédiatement rendue accessible au téléchargement. Aucun système de contrôle des informations, assuré par une ou plusieurs autorités expertes, ne sépare l'édition par un contributeur de l'utilisation par un autre individu. Dans un supposé contexte d'accroissement des *fake news*, auxquelles la presse et des groupes d'experts répondent par des services de *désintox*, **comment une source d'informations ne disposant pas d'un système de contrôle pourrait-elle inspirer la confiance d'utilisateurs potentiels ?**

L'utilisateur s'estime en confiance lorsqu'il juge sa source comme crédible. Dans le langage courant, la crédibilité est ce qui fait que quelqu'un ou quelque chose mérite d'être cru (*Le Petit Robert*, 2017). Elle dépend de la clarté et de la justesse logique du message, qui sont indispensables à une bonne réception des informations. Plusieurs sources de données géographiques sont reconnues crédibles par les professionnels du domaine (géomaticiens, enseignants-chercheurs, plateforme d'*open data* comme le site Internet data.gouv). Des instituts nationaux comme l'IGN en France, l'*Ordnance survey* au Royaume-Uni ou l'United States Geological Survey (USGS) aux États-Unis, sont considérés comme crédibles pour leurs expertises et la fiabilité de leurs services, qui garantissent une continuité dans la production de leurs informations. Elles font autorité dans le domaine de l'information géographique grâce à cette expertise mais également grâce à leur mission d'intérêt public. Ce rôle politique leur confère une légitimité. Depuis une quinzaine d'années, des producteurs commerciaux (TomTom, Here) ont également acquis des réputations semblables à celles des producteurs de référence. Ces producteurs de données sont des autorités légitimées par la

démonstration pratique de leur expertise[100]. Elles inspirent donc la confiance en tant qu'entreprises manifestement compétentes pour délivrer des informations de qualité et partagent une forme d'ancienneté et de continuité qui intervient comme critère de fiabilité. Elles ont fait preuve de régularité en matière de publications. La durabilité est ainsi clairement un facteur de fiabilité. La fiabilité d'une source est sa capacité à convaincre d'autres personnes qu'elle produit une information crédible. La fiabilité est donc un jugement qui se construit dans le temps et à l'usage. La disposition favorable du récepteur vis-à-vis de l'information ou de l'émetteur soutiendra la crédibilité du message reçu (Ginsca et *al.*, 2015, p.4). On peut estimer qu'un producteur « a fait ses preuve » lorsque ses services ont été jugés crédibles dans la durée. On lui fait alors confiance, en le créditant d'une capacité à ne publier que les informations qu'il a lui-même préalablement validées. On ne cherche pas systématiquement à vérifier les informations qu'il publie. Les producteurs peuvent axer leur communication sur cette garantie de fiabilité. C'est le cas de l'entreprise Michelin dont la marque de fabrique est d'assurer la sécurité de ses clients (fig. 18). La fiabilité de la carte contribue à cette sécurité.

Figure 18 Tweet présentant les cartes Michelin comme fiables, capture d'écran sur Twitter

Au contraire de ces autorités, les projets contributifs ne sont pas reconnus par les institutions de savoirs, universités ou instituts cartographiques, considérés comme des informations crédibles. Le producteur n'est pas ici une personne morale identifiée : en

[100] Les cartes et les GPS peuvent contenir des erreurs mais celles-ci doivent être suffisamment rares pour conserver la confiance de l'utilisateur. Par ailleurs, les producteurs nationaux sont associés, le plus souvent, à des cursus universitaires ou aux écoles d'ingénieurs, leur permettant d'affirmer le renouvellement méthodologique et la continuité de leur production. Les producteurs privés quant à eux peuvent mobiliser leur site internet afin d'exposer l'expertise de leurs équipes.

2019, OSM compte plusieurs centaines de milliers de contributeurs actifs[101]. Ce groupe est composé d'individus aux expertises très différentes et à l'engagement inégal. Dans les premières années du projet OpenStreetMap et d'autres projets contributifs en ligne, le contributeur a d'abord été considéré comme un simple amateur (Keen, 2007, Tapscott et Williams, 2006). En 2016, il est possible d'affirmer, au contraire, qu'au Royaume-Uni, en Allemagne et en France, la plupart des contributeurs présentent des attitudes que les amateurs ont rarement et montrent comportement professionnel (Yang et *al.*, 2016). L'expertise des contributeurs est devenue un nouvel élément de l'évaluation de la crédibilité de l'information géographique volontaire, par l'identification de comportements professionnels chez les contributeurs. Le professionnalisme n'est pas ici identifié par une situation d'employé rémunéré mais par des compétences techniques et théoriques et par une participation au projet soutenue dans le temps et en quantité. Ce comportement professionnel a été mis en lumière à partir de 2012 par l'identification d'un groupe de contributeurs très fortement impliqués en comparaison à l'ensemble des contributeurs (Neis et Zipf, 2012). Un des aspects récurrents du mode de production contributif a alors été exposé : un faible pourcentage de contributeurs produit l'essentiel des informations. De ce fait, il est devenu possible de mesurer l'expertise d'une plus petite partie de la communauté, celle des contributeurs « confirmés » (Neis et Zipf, 2012 ; Yang et *al.*, 2016). Ce groupe disparate de contributeurs n'inspire donc pas une confiance qui est le résultat «de routines, l'habitude rendant inutile un examen critique de la source [...] déléguée à des institutions » (Sahut, 2015). La nouveauté des projets libres exige que nous nous livrions « à une appréciation rationnelle » de la source (Sahut, 2015). La qualité des informations est le premier paramètre qui participe d'un jugement de crédibilité envers son producteur.

Le processus d'évaluation de la qualité de données qu'un producteur de données s'engage à suivre s'appuie sur des normes. Selon l'ISO, la qualité d'un objet dépend de sa propension intrinsèque à satisfaire des exigences[102]. Dans le domaine de

[101] À la date du 14 mars 2019, 5 258 283 membres sont enregistrés sur openstreetmap.org (Pascal Neis, https://osmstats.neis-one.org/?item=members). L'ensemble de ces contributeurs ne sont pas actifs et certains individus peuvent disposer de plusieurs comptes utilisés pour des imports, par exemple. L'activité des membres est en constante croissance par exemple, le 14 mars 2012 le nombre de contributeurs actifs était de 2837, le 14 mars 2015 le nombre de contributeurs actifs était de 2753 (un dimanche, jour de faible activité), le 14 mars 2018 il était de 4701 tandis que le 13 mars 2013 il était de 5071.

[102] La norme ISO 9000:2015(fr) intitulée Systèmes de management de la qualité — Principes essentiels et vocabulaire définit la qualité comme l'aptitude d'un ensemble de caractéristiques intrinsèques d'un objet à satisfaire

l'information géographique, ces exigences relèvent de la justesse logique, de la clarté et de la validité des informations géographiques en elles-mêmes et les unes par rapport aux autres, telles qu'intégrées dans une base de données. Les termes « clarté » et « validé » réunissent des paramètres classiques de la qualification d'informations géographiques qui reste un enjeu majeur de la cartographie, depuis ses origines. Des informations telles que les propriétés foncières, les distances ou les routes ont fait l'objet de recensements et de mesures pour ensuite être représentées en cartes. Ainsi, le problème de la qualité des informations a été rencontré dès les premières collectes de données. Dans ce chapitre, nous **traitons de ce problème de qualification des informations appliquées au cas des informations contributives produites dans le cadre du projet OpenStreetMap.**

Les analyses réalisées sur la base de données OpenStreetMap présentent une synthèse des facteurs de qualité dans la pratique de contribution. En effet, comme l'explique Amelot et *al.* (2015) : « La crédibilité renvoie à la fois au contexte de production et au degré d'ambiguïté de la donnée : plus le contexte de production est simple et plus l'ambiguïté autour de la source et des conditions d'acquisition sont faibles, plus la donnée peut être considérée comme crédible (Flanagin et Metzger, 2008)[103] ». La base de données OpenStreetMap est une base de données récente, au mode de production inédit. Son évaluation doit donc prendre en considération les directions et les tendances prises par les contributeurs afin d'envisager sa crédibilité à venir. Une source de connaissances n'est pas seulement « fiable » par ses résultats et ses acquisitions, elle l'est également par sa capacité à progresser en corrigeant ses erreurs et en se précisant. À travers l'évaluation de la « fiabilité » des données géographiques libres, on cherchera à apprécier leur caractère scientifique ou non.

Afin d'évaluer le processus de fiabilisation que connaît le projet OpenStreetMap nous procéderons en deux étapes. D'abord, nous présentons les mesures de qualité et de crédibilité des informations qui sont stockées en base de données (II.1. **Qualité et crédibilité des informations géographiques volontaires et libres**). Nous analyserons ensuite dans quelle mesure le mode de production de la base de données est scientifique

des exigences. Le terme « qualité » peut être utilisé avec des qualificatifs tels que médiocre, bon ou excellent. «Intrinsèque», par opposition à «attribué», signifie présent dans l'objet (3.6.1), consulté en ligne en janvier 2019 : https://www.iso.org/obp/ui/#iso:std:iso:9000:ed-4:v2:fr

[103] « *Two particularly important issues in this regard are the nature of gatekeeping in the digital media environment and the level of ambiguity surrounding both the source and the context of information*. » (Flanagin et Metzeger, 2008).

c’est à dire méthodique, ouvert à rectification et ceci dans un objectif d’objectivité (II.2. **Du projet OpenStreetMap à l’institution OpenStreetMap ?).**

2.1. Qualité et crédibilité des informations géographiques volontaires et libres

À partir d'un ensemble de publications scientifiques relatives à l'évaluation de la qualité d'informations géographiques volontaires, nous distinguons six paramètres classiques de l'évaluation de la qualité des données : l'exactitude de position, l'exactitude attributaire, la cohérence logique des objets les uns par rapport aux autres, l'exhaustivité, l'historique des données et l'actualité des données. Ces paramètres relèvent aujourd'hui de l'évaluation courante des données géographiques, apparaissant parfois sous d'autres terminologies. L'utilisation de données mais également l'évaluation de leur qualité repose d'abord sur la précision des métadonnées. Dans le cas d'une utilisation cartographique ou pour une évaluation, ces « données sur les données » peuvent être rédigées d'une manière littéraire ou être organisées selon un modèle standard. La date de production des données, la date de publication, le diffuseur, la licence de l'ensemble du jeu de données ou encore les méthodes de collecte sont autant d'éléments fondamentaux à connaître pour un bon usage des données concernées. Les métadonnées constituent ainsi le lien entre l'émetteur de l'information et le receveur. Il est donc fondamental que le cadre méthodologique soit transparent mais qu'il reste également accessible à la compréhension de l'utilisateur le plus probable. En effet, la bonne transmission des conditions de production apporte à l'utilisateur la connaissance de la concordance de son usage avec l'ensemble de données. En outre, certaines métadonnées contiendront directement des informations relatives à la qualité des données comme les mesures relatives à la position des attributs.

En 2009, durant le Symposium International sur la qualité des données spatiales, six chercheurs spécialistes de l'évaluation de la qualité des données spatiales rappellent, en préambule d'une synthèse des travaux réalisés sur la qualité des données spatiales de 1980 à 2010, que des premiers travaux de statistique ou de métrologie ont été mené afin de proposer des mesures d'exactitude de position des objets géographiques et ce dès les années 1950 (Devillers et *al.*, 2010). Pepijn van Oort (Van Oort, 2006) propose un état de l'art détaillé relatif aux recherches sur la qualité des données géographiques dans son travail de thèse. Il date l'accroissement des publications sur cette question aux années 1960, années de l'apparition des SIG. Il situe aux années 1970 l'apparition de l'usage d'images satellites pour la production d'informations géographiques. Ces deux changements dans la pratique de la cartographie et dans l'étude de l'information

géographique ont entraîné un fort accroissement de la quantité d'informations produites. Il n'est alors plus possible de les qualifier manuellement. Cependant, ce n'est qu'à partir des années 1980 durant lesquelles la communauté scientifique se saisit de la question, intégrant directement à la réflexion, sur la fiabilité des données produites, le concept de qualité des données dans le mouvement de la définition des standards de données spatiales (Devillers et *al.*, 2010). Ainsi, des auteurs comme Nicholas Chrisman (1991) et Michael Goodchild et Sucharita Gopal (1989), intègrent la pratique de l'évaluation des données, pour un bon usage de celles-ci, dans leurs définitions des activités fondamentales du domaine des SIG. Van Oort signale également le travail d'Alan Brimicombe (2003) qui montra, en la matière, une activité de recherche intense, des années 1990 au début 2000. Le nombre de publications s'accroit alors significativement, ainsi que le déroulement de conférences nationales et internationales consacrées à la qualité des données spatiales[104].

À la fin des années 1980, les nouveaux standards de mesure de la qualité des données géographiques reposent sur cinq « dimensions fondamentales » : l'exactitude de position, l'exactitude attributaire, la cohérence logique, l'exhaustivité et l'historique des données[105]. Ceux-ci ont été admis par un consensus, permettant la rédaction du *Content Standard for Digital Geospatial Metadata*[106] (Goodchild et Li, 2012). En 1995, les chercheurs Gutpill et Morrison ajoutent les « dimensions » de l'actualité des données et de l'exactitude sémantique (Guptill et Morrison, 1995 ; Goodchild et Li, 2012)[107].

Considérons chacun des six paramètres abordés dans cette section afin de nous appuyer sur les procédures traditionnelles d'évaluation des données géographiques qui ont pu être appliquées à la base de données OSM. L'analyse de **l'historique des données** (1) s'appuie sur la source à partir de laquelle les données ont été produites. Dans le cas où une donnée est produite suite à l'interprétation d'une autre donnée, la qualité de la

[104] Pepijn Van Oort cite l'ouverture du *Symposia on Spatial Accuracy Assessment* en 1996 et du Symposia on Spatial Data Quality en 1999. Il cite également des publications comme celles issues des conférences citées ci-dessus et le chapitre de Nicholas Chrisman dans la première édition du manuel de SIG de référence *Geographical information systems* dirigé par David Maguire, Michael Goodchild et David Rhind. Nous ajoutons que la deuxième édition, de 2005, de ce manuel dédie une plus large partie encore à la question de la qualité des données.

[105] « *In the 1980's discussions over a potential geospatial data standard for the US Federal Government led to a consensus on five fundamental dimensions: positional accuracy, attribute accuracy, logical consistency, completeness, and lineage* » (Goodchild et Li, 2012).

[106] Cette norme de rédaction de métadonnées à l'échelle des États-Unis vise, dès 1990, à soutenir les partenariats entre les départements de l'agriculture, du commerce, des transports de la NASA et de l'urbanisme, consulté en ligne en avril 2019 : https://www.fgdc.gov/metadata/csdgm/

[107] « *More recent literature has suggested the addition of further elements, including temporal accuracy and semantic accuracy* (Guptill and Morrison, 1995) » (Goodchild et Li, 2012).

nouvelle donnée dépend de celle dont est issue l'interprétation. Mais elle dépend aussi de la méthode d'interprétation d'autre part. Dans le cas OpenStreetMap, c'est du renseignement de la source par le contributeur pour le *changeset* ou pour chaque objet que dépend la traçabilité de l'historique de la donnée. Une information fiable et juste doit correspondre à l'état actuel des connaissances ou des observations de ce qu'elle décrit. Un des paramètres de l'évaluation de la qualité des données vise ainsi à assurer **l'actualité des données** (2). Le plus souvent, un fournisseur de données publie un jeu de données actualisé périodiquement et annonce les mises à jour qui ont été faites dans les métadonnées de la nouvelle publication. Pour le cas OpenStreetMap l'actualité des données n'est pas annoncée périodiquement. Cependant, à chaque contribution d'une information sur un objet, la date de cette édition est enregistrée. C'est un premier élément indicateur d'une actualisation de la donnée si l'édition est récente ou d'un faible maintien de l'information, dans le cas inverse.

Par la mesure de **l'exactitude de position** (3) et **des attributs** (4), on cherche à évaluer la conformité à la réalité. La conformité à la réalité n'est par définition pas quantifiable, puisque « la valeur vraie »[108] est, en pratique, impossible à connaître. Ainsi, l'exactitude de position et l'exactitude d'attribut s'évaluent souvent relativement à un modèle de référence (Hacklay, 2010 ; Girres et Touya, 2010 ; Zielstra et Zipf, 2010). Ce procédé d'évaluation correspond à la définition des termes relatifs à la qualité des données géographiques par l'ISO selon laquelle l'exactitude est la « précision, justesse de l'accord entre un résultat de test ou mesure et la valeur vraie [...]. La vraie valeur peut être une valeur de référence acceptée comme vraie »[109]. Il existe un flou terminologique entre « exactitude sémantique » et « exactitude d'attributs ». Les deux formules sont parfois confondues en une même définition comme la correspondance entre la réalité (par exemple, une autoroute) et les attributs qui la décrivent (une voie de circulation, large, réservée aux véhicules automobiles, sans croisement, à chaussées séparées destinées chacune à un sens de circulation). On mesure l'exactitude d'attributs en fonction de la correspondance des termes de la description avec la réalité. En revanche, l'exactitude sémantique renvoie à la signification, c'est-à-dire à l'interprétation possible du récepteur, et aux attributs choisis par l'émetteur au moment de la définition

108 Vocabulaire international de métrologie – Concepts fondamentaux et généraux et termes associés (VIM) 3e édition, 2008, https://www.bipm.org/utils/common/documents/jcgm/JCGM_200_2012.pdf ; consulté en ligne en mars 2019

109 ISO 19157:2013(fr) Information géographique — Qualité des données; consulté en ligne : https://www.iso.org/obp/ui/#iso:std:iso:19157:ed-1:v1:fr

du modèle de données. Cette troisième exactitude évalue donc davantage la qualité du modèle de données conçu traditionnellement à la création de la base de données et en fonction d'un besoin que la qualité de la donnée produite finalement. Lorsque l'on s'intéresse à la qualité sémantique, dans son sens unique et non confondu avec la qualité attributaire, on se demandera si la donnée est interprétable par l'utilisateur. Cette interprétation est un des éléments de la crédibilité d'un message, il doit être compris sans ambiguïté. La communication entre le producteur et l'utilisateur dépend de la qualité de la documentation accompagnant les métadonnées, qui doit être comprise en dehors du milieu expert. En outre, la tendance au partage et au croisement des informations de sources différentes accroît ce besoin en métadonnées (Devilliers et al, 2010).

Un autre paramètre d'évaluation de la qualité est **la cohérence logique** des objets les uns par rapport aux autres (5). On cherche à mesurer les liens topologiques, de superposition et d'inclusion des objets les uns par rapports aux autres. Il s'agit d'un paramètre propre aux objets géographiques dont le caractère géographique lui-même repose largement sur les relations physiques qu'il partage avec son environnement. La cohérence logique des données OSM est évaluée dès les premières recherches relatives à la qualité des données de la base OpenStreetMap. Par exemple, des erreurs de connexité dans les réseaux routiers sont relevées dans les travaux de recherche évaluant la données (Girres, Touya, 2010 ; Neis et *al.*, 2012). Ce défaut limite fortement l'utilisation d'un jeu de données issu d'OSM dans les domaines d'applications comme le calcul d'itinéraire sans traitements géomatiques lourds. En 2010, en France l'absence de superposition entre une frontière et l'objet physique qui la constitue parfois (comme une route) est une erreur de cohérence logique très présente (Girres, Touya, 2010). La superposition d'objets physiques et conceptuels comme les frontières sont indispensables à une représentation cartographique de qualité. Le cartographe aura tendance à décaler le cours d'eau ou la route pour le confort de lecture, mais ce décalage doit être volontaire et maîtrisé. Chacun de ces défauts ont fait l'objet de campagnes d'amélioration[110], par des contributeurs fortement engagés dans l'amélioration des données, dans le cadre des préconisations de contribution décrites dans le *wiki* OSM.

[110] En France, il est possible de retracer la cartographie couverture complète des limites administratives du territoire sur la liste de discussion Talk-fr : https://lists.openstreetmap.org/pipermail/talk-fr/2013-December/subject.html
En France encore, les contributeurs locaux proposent régulièrement des « projets du mois » qui consistent en de simples objectifs à réaliser collectivement en un mois comme par exemple atteindre l'exhaustivité en matière de

Ces caractéristiques d'exactitude et de cohérence logique doivent être reconnues sur un ensemble de données théoriquement exhaustif pour des thématiques précises et selon un territoire. En mesurant **l'exhaustivité** (6), il est possible de repérer une absence de données et/ou une présence excédentaire (Van Oort, 2006). Une information est plus intéressante lorsqu'elle est complète. En effet, un défaut d'information est, par définition, impossible à connaître sans une étude consacrée à son identification. Pour un utilisateur, l'information doit donc être complète pour la zone qui concerne son usage. On parle de collecte systématique. Cette collecte consiste à parcourir l'ensemble d'une zone sans *a priori* de localisation des objets à relever. La collecte permet de localiser chacun des objets d'une catégorie. Cette catégorie est décrite par un ou plusieurs attributs. Pour une bonne qualité d'exhaustivité attributaire, les attributs spécifiés comme nécessaires à la définition d'un objet doivent donc être renseignés. Par exemple, les sens uniques sur les voies de communication sont des attributs nécessaires à l'utilisation des données pour du calcul d'itinéraire. Afin d'évaluer correctement l'exhaustivité, il est important de se concentrer sur les éléments prescrits par les spécifications de production. Il n'est pas attendu de toutes les bases de données d'intégrer les chemins de randonnées ou le nombre d'habitants par unité administrative relevé au dernier recensement. Selon notre usage de la définition de l'exhaustivité, l'absence d'un type d'objet dans les spécifications du modèle de données ne relève donc pas d'un défaut d'exhaustivité.

L'exhaustivité de couverture d'une thématique peut être mesurée en fonction d'un jeu de données de référence. L'évaluation est donc relative (Zielstra et Zipf, 2010) et le jeu de données est évalué en fonction d'un autre jeu de données et non en fonction du réseau routier réel. L'exhaustivité de couverture n'est pas évaluable pour l'ensemble de la base OSM, qui a pour emprise la surface terrestre mondiale. L'exhaustivité est toujours réduite à un extrait de la base. En 2010, Haklay propose de mesurer l'exhaustivité d'un extrait de la base de 2008 découpée autour de l'Angleterre (Haklay, 2010a). Il applique une méthode par sondages sur de petites zones dans l'emprise du pays et d'une plus large autour de Londres. Haklay calcul la longueur totale des routes par zones. Il mesure ensuite les différences en comparant ce premier calcul à la longueur totale des routes du jeu de données de référence découpé selon les mêmes zones. Lorsqu'apparaît une

postes de police et de gendarmerie, de panneaux électoraux ou encore des postes électriques en s'appuyant des jeux de données ouverts par les administrations.

différence entre les longueurs globales, cela signifie que l'un des jeux de données est plus exhaustif que l'autre (Haklay 2010, Zielstra et Zipf, 2010). Une mesure plus originale est celle proposée par Keßler et De Groot (2013). Ceux-ci choisissent de se concentrer, pour mesurer l'exhaustivité attributaire, sur la mention des spécificités. Ils n'évaluent pas les *tags* évidents, les *tags* principaux comme `highway=*` indiquant une route ou `name=*` précisant le nom de l'objet mais plutôt un ensemble de « *tags* secondaires obligatoires ».

Une fois ces six paramètres classiques énumérés, on commence à percevoir une des principales difficultés auxquelles les chercheurs ont fait face dans les premiers moments de leurs évaluations de la qualité des données géographiques volontaires et libres. L'exactitude de position, l'exactitude attributaire, la cohérence logique et l'exhaustivité sont des paramètres fortement dépendants du modèle de données. Celui-ci est lui-même fortement dépendant du besoin auquel la base de données est conçue pour répondre. Ainsi, des mesures, exactes, à moins d'un mètre, seront indispensables à un projet d'aménagement de la voirie. Un réseau routier modélisé en graphe est indispensable à la diffusion d'un service de navigation routière ou piétonne. Cependant, les contributeurs au projet OpenStreetMap défendent que la base de données OSM n'est pas conçue pour un usage spécifique et qu'aucun modèle de données figé n'est imposé. OpenStreetMap ne s'engage pas à respecter les standards de production partagés par les acteurs traditionnels. OpenStreetMap n'est pas non plus une formation organisée et hiérarchisée mais davantage un groupe de personnes aux intérêts divers et aux principes partagés. De ce fait, la base de données du projet OpenStreetMap n'est pas enrichie quotidiennement selon un même ensemble de méthodes de collecte et de modélisation. Des contraintes techniques évitent l'ajout de géométries incohérentes dans certains logiciels d'édition de la base de données. Par ailleurs, l'actualité des données n'est pas garantie par la publication régulière de données mises à jour. Elle est continue et l'absence d'édition des données ne signifie pas qu'elles sont invalides. En effet, l'absence de nouvelle contribution sur un objet ne signifie pas que les informations qu'il porte sont obsolètes pour autant.

L'exhaustivité en information est, en principe, l'une des opportunités, avec l'actualité des informations, offerte par la production volontaire d'informations. Cependant, il faudrait pour cela atteindre une couverture complète d'un territoire en termes de

données et de répartition des contributeurs sur le territoire qui permettrait aux contributeurs de maintenir à jour un jeu exhaustif de données. En outre, cette idée n'est pas cohérente avec le projet OpenStreetMap qui propose plusieurs centaines de thématiques, les principaux *tags*, dont la visée est de couvrir l'ensemble des espaces terrestres. Ainsi, dans la mesure où les contributeurs collectent des informations relatives à des thématiques qui les intéressent, sans coordination hiérarchisée, on ne peut s'attendre à une couverture systématique de plusieurs centaines de *tags* dans le monde. Néanmoins, dans le but de rendre la base de données opérationnelle, des contributeurs se sont occasionnellement organisés afin de couvrir un territoire par le renseignement de son découpage administratif. D'autres projets visent à organiser les contributeurs autour d'une thématique et à leur proposer des outils de contribution simple dans le but de couvrir davantage de territoires tout en utilisant les mêmes spécifications.

Dans les années 1990, le concept d'incertitude, propre à la métrologie, a été intégré aux questions d'évaluation de la qualité des données. En intégrant ce terme, on cherche d'abord à améliorer l'évaluation de la qualité de position des objets. L'incertitude est conçue pour admettre la possibilité d'erreurs (donc un intervalle à l'intérieur duquel se trouvent la valeur exacte, inconnue, et la valeur calculée d'une grandeur), d'imprécision (due à une documentation et des définitions insuffisantes) et d'ambigüité dans la définition des objets (Fisher, 1999). Cependant, le concept d'incertitude peut être appliqué à chacun des paramètres d'évaluation de la qualité. Lorsque Goodchild se saisit de la question de l'incertitude, c'est pour alerter sur l'illusion pour l'information géographique sous forme numérique, d'être au plus proche de la réalité, alors qu'elle correspond toujours à une modélisation[111] (Zhang et Goodchilld, 2002 ; Goodchild et Li, 2012). La traduction d'observations en informations géographiques, en modèle de bases de données géographiques et en cartes sont des productions issues d'un travail d'abstraction et de simplification de la réalité. Les inexactitudes et les incertitudes font partie de ces représentations géographiques. En outre, celles produites à l'origine peuvent être propagées lors de nouvelles opérations de production à partir de ces représentations. L'incertitude initiale doit donc être connue d'un utilisateur qui ne serait

111 « In practice, however, geographic information systems (GIS) rarely keep track of measurement accuracy, preferring to process and report coordinates as if they were perfect » (Goodchild et Li, 2012).

pas le producteur de l'information. C'est le rôle de l'historique des données et plus largement des métadonnées.

Ces six paramètres d'évaluation sont à relativiser. Compte tenu des limitations en ressources financières que connaissent les producteurs de données, il s'agit d'évaluer la qualité de manière raisonnable. Il faudra ainsi parfois accepter l'absence de la mention de la source primaire dans les métadonnées ou un retard dans la mise à jour de jeux de données. La qualité repose essentiellement sur la rareté de ces défauts ainsi que sur la capacité du système d'exploitation de la base de données pour les identifier.

2.1.1. Comparer OSM à une référence

Une base de données de référence est une base de données qui est certifiée par une institution d'autorité. Lorsque son autorité lui confère une crédibilité du fait de ses capacités de maintien du service, de la fiabilité de son information et de son expertise, une base de données de référence peut constituer un point de comparaison afin d'évaluer la qualité d'une base à tester. Ainsi, en 2010, Haklay décrit-il l'agence cartographique du Royaume-Uni, l'*Ordnance Survey*, comme une agence connue pour son système de contrôle de qualité (Haklay, 2010a). Il affirme alors que le jeu de données de l'*Ordnance Survey* représente, du fait de son autorité, une plus grande précision que la base de données OSM et une qualité globale incontestable, au moins en ce qui concerne la position et les attributs de chaque objet (Haklay, 2010). Cependant, il modère son propos en commençant par rappeler la nouveauté du projet OpenStreetMap qui, compte tenu de son mode de production, ne peut déjà avoir égalé l'ampleur des bases de données proposées par l'*Ordnance Survey*. Il pointe donc directement l'importance de prise de recul et d'historicité dans la qualité d'une base de données.

Des bases de données commerciales peuvent également être choisies comme bases de données de référence. Zielstra et Zipf avancent que le succès de la base *Multinet* de Tele Atlas[112] suffit à faire d'elle une référence (Zielstra, Zipf, 2010). Choisir *Multinet* comme référence est conforme aux travaux antérieurs sur la qualité des données

112 Propriété de TomTom depuis 2008, la base de données est utilisée pour les services cartographiques en ligne comme ViaMichelin et Mappy (notamment pour les petites et moyennes échelles).

géographiques volontaires. Les auteurs signalent néanmoins que la base *Multinet* doit être considérée comme une référence pour évaluer la qualité de données en fonction de l'usage pour laquelle la base est conçue. Par exemple, si on souhaite évaluer un ensemble de données qui sera utilisé pour du calcul d'itinéraire, il est approprié de choisir des bases de données commercialisées par TeleAtlas (TomTom depuis 2008).

Une base de données de référence, qui correspond à un pays ou à un continent, ne permet en aucun cas d'évaluer l'ensemble de la base de données OSM qui présente une emprise mondiale. Cependant, cette méthode a démontré son efficacité par sondages. D'abord, dès 2009, il a été possible d'évaluer les contributions dans les pays qui se sont rapidement saisis du projet OSM comme l'Allemagne, le Royaume-Uni et la France. Ensuite, une base de données de référence sert également localement à questionner les choix de modèles de données par la connaissance des spécificités locales qui ont consolidé la définition de son modèle de données. Enfin, les comparaisons à des bases de données de référence ont permis aux chercheurs de mettre en évidence des facteurs d'amélioration de la qualité qui sont internes au projet. L'analyse de ces facteurs nécessitaient davantage de matériaux, le développement de certaines fonctionnalités arrivées plus tard dans le projet (comme la publication de l'historique et l'invention des *changesets*) et les résultats des premiers travaux de comparaison à des bases de données de référence dont le mode de production étaient connues par les chercheurs. Nous précisons que cette méthode de comparaison à une référence reste d'actualité pour évaluer les propositions d'évaluation s'appuyant sur les propriétés intrinsèques des objets.

La méthode généraliste par comparaison à une référence peut s'appliquer à la question de l'exactitude de position. Nous avons cité les travaux d'Haklay sur les tests exhaustivité (Haklay, 2010). En 2010, en cherchant lui aussi à évaluer la qualité de position, il compare l'Angleterre, telle que représentée en 2008 dans OSM, à des jeux de données de référence produits par l'Ordnance Survey. Dans la région de Londres, Haklay mesure les écarts entre la géométrie du réseau routier d'OSM et les objets de référence de l'Ordnance Survey. Il conclue à un écart moyen de six mètres entre ces deux sources[113], qu'il juge acceptable[114] dans un but cartographique. Dans un autre

113 Il est important de rappeler qu'un GPS grand public ne peut dépasser une précision située entre 6 et 10 mètres.
114 « *reasonable accuracy* » (Haklay, 2010) ; Comme pour chaque évaluation, cette précision est ici jugée acceptable compte tenu de l'usage de la base de données.

article, il poursuit son évaluation de la donnée OSM et reprenant ses précédents sondages, il conclue à une corrélation entre le nombre de contributeurs et la qualité de position (Haklay et *al.*, 2010). À partir de plus de quinze contributeurs par km², l'écart de précision de position est inférieur de 6 mètres par rapport à la base *Integrated Transport Network* de l'Ordnance Survey. Il ajoute une observation à ses conclusions, souvent vérifiée par d'autres chercheurs : les cinq premiers contributeurs dans une zone « semblent fournir la plus grande contribution en termes d'amélioration de la précision de position » (Haklay et *al.*, 2010, p.321). En 2012, en Allemagne, Neis, Zielstra et Zipf considèrent que selon une tendance qu'ils avaient révélée, plus de six contributeurs sont nécessaires pour constituer un jeu de données routier « quasi complet » par kilomètre carré (Neis *et al*. 2012)[115].

La comparaison entre la base OSM et des bases de référence permet d'obtenir les premiers résultats qui présentent des écarts moyens « acceptables » concernant l'exactitude de position du réseau routier OSM (Haklay, 2010 ; Girres et Touya, 2010 ; Neis et *al.*, 2012). Cependant, les chercheurs soulignent la forte dispersion statistique des écarts, révélatrice d'une forte hétérogénéité dans l'exactitude de position[116]. Le nombre variant de contributeurs sur une zone ou par objet a rapidement été considéré comme facteur principal de cette hétérogénéité. Girres et Touya (2010), de même que Neis, Zielstra et Zipf (2012), posent l'hypothèse qu'à un nombre important de contributeurs est associé un faible écart entre le jeu de données testé et le jeu de données de référence. Faut-il alors chercher plus loin et mettre en place des méthodes spécifiques de collecte pour garantir la fiabilité de la base dans ce domaine ?

Afin de présenter des garanties et de produire ce sentiment de confiance, les producteurs de données de référence ont mis en place d'autres méthodes, notamment pour la collecte et le dessin (lecture d'images, généralisation) des objets géographiques qu'ils détaillent à leurs utilisateurs à travers des documentations précises. Ces documentations spécifient le modèle de données, les contraintes d'intégrité imposées en cas de tentatives de modification ou encore les moyens mobilisés pour effectuer la collecte. Un producteur de référence garantit le respect des contraintes et des spécifications

115 « *Bearing in mind with the current data collection trend in Germany, completeness for car navigation data of more than 90% could already be achieved in relative comparison to the commercial dataset with an average of two project participants per square kilometer. According to the trend line, more than six participants are required to achieve a dataset that is close to "complete*" » (Neis et *al.* 2012).

116 « *The distances computed between OSM and BT TOPO® road networks are close to reference accuracy but the standard deviation nearly as important as the means shows heterogeneity* » (Girres et Touya, 2010).

décrites dans sa documentation. Ainsi, le fait de comparer un certain jeu de données avec une base de référence peut mettre en évidence un manque de spécification dans le mode de production des données évaluées. C'est ce que montre la comparaison entre le tracé de côte de la France en 2010 et la BD TOPO de l'IGN (Girres et Touya, 2010). Girres et Touya remarquent que cet exemple « illustre l'importance de définir des spécifications précises pour éviter la subjectivité » des contributeurs. Par ailleurs, ces auteurs remarquent une différence de qualité entre les différents niveaux hiérarchiques des attributs de la thématique routière. Près de 100 % des routes principales présentent le même niveau hiérarchique avec leur « homologues » du jeu de données de référence composé d'extrait de la Base de données TOPO. À partir du niveau hiérarchique des routes secondaires, ce résultat chute à 49 % de correspondance avec le jeu de données de référence. Les auteurs expliquent cet écart par une sous-estimation des voies secondaires décrites comme des voies résidentielles par les contributeurs à OSM (Girres et Touya, 2010). En 2012, l'étude de Neis, Zielstra et Zipf présentent des résultats similaires en Allemagne : en 2008, le réseau autoroutier allemand apparaît comme complet dans OSMet en 2009, le réseau secondaire l'est également. En 2010, date de fin de leur analyse, les réseaux résidentiels restent alors incomplets mais leurs densités sont croissantes.

Lors de la comparaison à une base de données de référence, l'exhaustivité des données est le deuxième paramètre de la qualité qui est le plus régulièrement évalué, en même temps que celui de l'exactitude de position. En 2010, selon Haklay (2010), les réseaux routiers des centres des grandes villes (Londres, Manchester, Birmingham, Newcastle et Liverpool) sont de longueur comparable aux extraits de référence et la couverture de ces zones est homogène. L'exhaustivité présente cependant une forte diminution depuis les zones urbaines jusqu'aux zones rurales. Haklay (2010) mesure également une diminution encore plus prononcée de l'exactitude attributaire. Les voies en zones qu'il qualifie de rurales, sont rarement nommées. Le chercheur propose deux facteurs explicatifs à cette faible exhaustivité attributaire. D'une part, il rappelle l'accès limité à des cartes libres de droit comme sources d'informations pour la contribution. D'autre part, d'après lui, les contributeurs urbains des centres-villes ne pratiquent pas l'enquête de terrain dans des zones peu attractives à leurs yeux. On comprend alors l'importance des images satellites comme facteur favorable à la couverture d'un territoire, bien que leur exploitation reste limitée à l'identification de la géométrie des objets et à quelques

déductions thématiques comme l'occupation du sol ou la présence de bâtiments. En effet, Haklay (2010) souligne le lien entre l'accès aux images satellites haute définition de Yahoo sur Londres et un dessin qui présentait, en 2010, une exactitude de positon plus élevée que dans d'autres régions de l'Angleterre.

L'exhaustivité peut également s'évaluer en comparant un extrait d'OSM à une base de données de référence commerciale comme la base de données *MutliNet* de TeleAtlas (Zielstra et Zipf, 2010). En 2009, en Allemagne, la longueur des rues disponibles dans OSM est inférieure à celles disponibles dans *MultiNet*. Cependant, le taux de croissance de cette mesure de la longueur de l'ensemble du réseau de rues dans OSM est déjà impressionnant. En effet, d'avril à décembre 2009, la différence kilométrique du réseau OSM avec celui de la base *Multinet* est passé de 29 % à 7 % (Zielstra et Zipf, 2010, fig. 19).

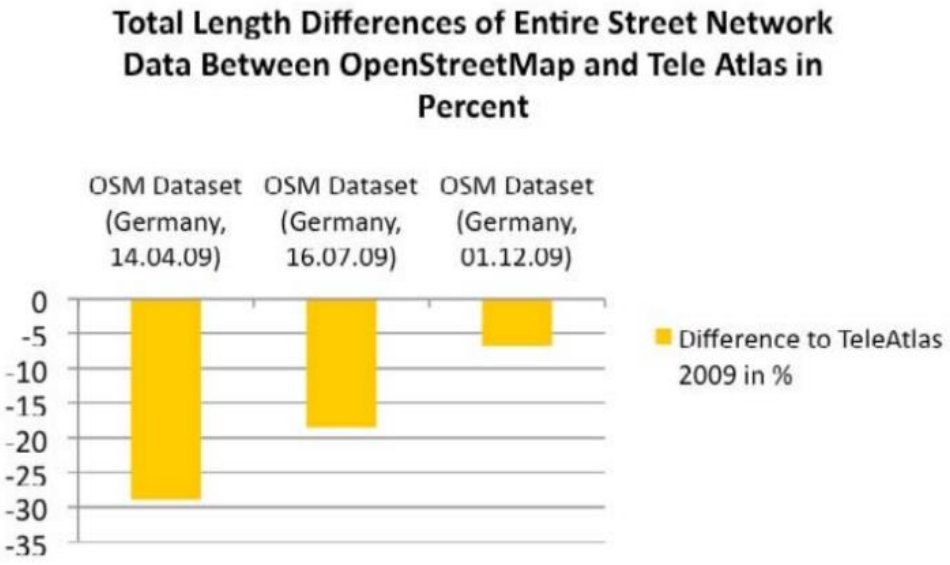

Figure 19 Tendance à un rapprochement de la densité des réseaux OSM et Mulinet par Denis Zielstra et Alexander Zipf (2010)

L'évaluation par comparaison à un ensemble d'informations de référence permet donc d'analyser de manière simple et selon des méthodes éprouvées l'exactitude de position, l'exactitude d'attribut et une certaine exhaustivité.

Les comparaisons, discutées ci-dessus, permettent de conclure que les jeux de données étudiés issus d'OSM sont en général « comparables » aux données géographiques traditionnelles qui sont maintenues par les agences nationales de cartographie et les fournisseurs commerciaux. Cependant, ces conclusions encourageantes sont

relativisées par l'identification de fortes hétérogénéités dans la mesure de ces paramètres[117]. Selon certains chercheurs, cette hétérogénéité rendrait la base de données OpenStreetMap inutilisable, en l'état, pour des usages très spécifiques. Par ailleurs, comme l'indique Barron et. *al.* (2014), les bases de données de référence propriétaires ne sont accessibles qu'à des coûts élevés et selon des licences restrictives. De ce fait, des analyses alternatives ont été conçues et l'évaluation des données par leurs seules caractéristiques intrinsèques est une méthode largement plébiscitée par la recherche.

L'évaluation par comparaison à un jeu de données de référence est une méthode qui permet d'identifier des tendances de contribution menant à des niveaux de qualité différents. Dans certains travaux exploratoires, on a inféré des facteurs explicatifs d'une qualité jugée satisfaisante selon les propriétés intrinsèques des objets. Selon les recherches que nous allons passer en revue, des seuils ou des typologies permettent de juger la crédibilité des données sans recourir à une base de données de référence, une fois que les critères sont jugés satisfaisants.

2.1.2. Qualité d'une base de données en fonction des propriétés intrinsèques des données

Les propriétés dites intrinsèques sont issues de la nature propre des objets de la base de données. Elles constituent leur essence et caractérisent donc leurs qualités intrinsèques. Le contributeur d'un objet inclut, dans la base de données, le nombre de nœuds qui le constituent, le nombre d'éditions de cet objet, les types de transformations issues de ces éditions, la diversité de ses attributs, l'ancienneté de la dernière contribution : chacun de ces éléments constitue la nature d'un objet. La mesure de la pertinence d'un objet est alors aussi fonction de ses caractéristiques intrinsèques. Du fait de la variété et de la quantité des paramètres mesurés, les méthodes d'évaluation de la qualité intrinsèque se distinguent donc de la méthode de comparaison et de référence qui est davantage normée. Ainsi, étudier les caractéristiques intrinsèques des données OpenStreetMap permet de considérer la contribution dans toute son originalité. Les caractéristiques intrinsèques, citées ci-dessus, en sont les exemples mobilisés le plus couramment.

117 Précisons que, majoritairement, les études ont été conduites sur des territoires où les communautés locales sont particulièrement présentes et souffrent donc plus faiblement du facteur d'hétérogénéité.

Les premiers travaux sur la qualité intrinsèque ont été inspirés par la « Loi Linus » (Haklay et *al.*, 2010 ; Goodchild et Li, 2012 ; Keßler, et De Groot, 2013)[118]. La mise en évidence d'un processus d'accumulation des informations par itérations pour la description des objets a mené les chercheurs à s'intéresser aux historiques des objets afin d'en évaluer la qualité. En 2014, Barron et *al.* (2014) s'appuient sur l'historique des données OSM pour identifier des tendances de contribution, indiquant une constance de la part d'une communauté et des états de stagnation, marquant la complétude d'une thématique ou l'essoufflement de la contribution. Les auteurs proposent l'outil iOSMAnalyzer qui comprend un ensemble de vingt-cinq indicateurs d'évaluation et indicateurs de qualité des données. Le jeu de données comprenant l'historique des créations, modifications et suppressions est ainsi produit depuis octobre 2007. Considérant que la donnée doit être évaluée en fonction de l'usage qui en sera fait, les auteurs scindent leurs vingt-cinq indicateurs en six catégories : « informations générales », « géocodage », « points d'intérêts », « calcul d'itinéraires », « applications cartographiques » et « comportement des utilisateurs » (fig. 20).

118 La loi Linus est expliquée dans la sous partie 1.2.2 « Volontaire, contributif ».

Figure 20 Résumé des 25 indicateurs proposés par Barron et al. (2014), figure intitulée "Overview of the iOSMAnalyzer's intrinsic Quality Indicators" par les auteurs

Chacune de leurs analyses sont concentrées sur une zone et non sur l'ensemble de la base de données OSM. Pour la première catégorie, la plus généraliste et peu tournée vers un usage des données, les auteurs s'intéressent à l'accroissement quantitatif du jeu de données sur la zone étudiée, à l'évolution de la qualification des objets à l'aide de davantage de *tags*, à l'actualité des données et de l'exactitude de position. Cette exactitude de position est évaluée en comparant la position actuelle de l'objet à sa précédente et ceci pour l'ensemble du jeu de données. En étudiant les intersections de routes, il est possible de noter une tendance à la correction d'un potentiel décalage issu du mauvais géo-référencement de l'image satellite utilisée auparavant. En ce qui concerne la catégorie des points d'intérêts (POI), les auteurs soulignent qu'une simple augmentation quantitative de leur nombre est en général un indicateur positif. Ensuite, ils considèrent qu'une augmentation du nombre *tags* par POI montre une tendance vers une description plus proche de la réalité. Enfin, à la manière de Keßler, et De Groot (2013), ils constituent une liste de clés indispensables à la description de POI et l'emploient pour évaluer la complétude attributaire de chaque objet.

En 2013, Keßler et De Groot ont proposé un indicateur de confiance à attribuer à chaque objet de la base OSM. Le résultat est une classification en cinq classes d'informations, des moins fiables aux plus fiables. Cet indicateur est composé de cinq variables, d'égale importance, qui donnent une évaluation des propriétés intrinsèques des objets. Les propriétés choisies sont les suivantes : pour un même objet, un nombre élevé de contributeurs, un nombre élevé de versions et une tendance forte à la contribution dans une zone, indiquant une confirmation indirecte des objets de cette zone, sont considérés comme des variables positives. Le nombre de contributeurs et de versions différentes sont des paramètres proches de la « Loi Linus » et du principe d'itération propre au développement de projets contributifs. Plus précisément, on considère qu'un élément est amélioré s'il est revu et éventuellement corrigé par plusieurs individus différents. En conclusion, Keßler et De Groot évoquent la réputation des contributeurs comme pouvant être un paramètre supplémentaire de l'évaluation de la confiance à accorder à l'information portée par un objet de la base.

Afin de tester leur indicateur, Keßler et De Groot commencent par évaluer empiriquement 74 objets de la base de données qui correspondent à un quartier de la ville de Munster en Allemagne. Ils procèdent à une enquête de terrain évaluant la qualité des données selon leurs observations. L'évaluation qu'ils choisissent pour valider leur indicateur consiste à mesurer l'exactitude attributaire, la cohérence topologique et l'exhaustivité attributaire des 74 objets. Leur enquête de terrain leur permet de vérifier la validité du *tag* principal de chaque objet (comme `amenity=restaurant`) mais également des *tags* secondaires comme l'adresse pour un POI ou le nombre de voies pour une route. Keßler et De Groot concluent que 77 % des informations sont parfaitement correctes (57 objets). La position des objets les uns par rapport aux autres (cohérence topologique) a également été vérifiée pendant l'enquête de terrain. Un seul objet parmi les 74 objets est alors placé du mauvais côté de la rue. Enfin, ils évaluent l'exhaustivité attributaire des *tags* secondaires jugés obligatoires. Toujours dans le cadre de leurs observations, 57 % des objets correspondent pleinement aux exigences de qualité. Lorsqu'ils excluent le paramètre d'exhaustivité attributaire des *tags* secondaires, 81 % des objets sont de très haute qualité. À cette enquête de terrain, Keßler et De Groot comparent les résultats de l'indicateur, appliquées aux mêmes 74 objets. L'indicateur estime l'ensemble des données modérément fiable. Un petit nombre d'objets atteint le plus haut niveau de fiabilité. De manière générale, l'indicateur de

confiance qu'ils ont conçu sous-estime la qualité des données par rapport à leurs estimations par leur enquête de terrain. Cependant, l'indicateur et l'enquête de qualité présentent les mêmes tendances. Grâce à l'enquête, on sait que des caractéristiques intrinsèques permettent de juger de la fiabilité des objets analysés. Nous comprenons le choix des auteurs de se concentrer sur des *tags* spécifiques aux objets modélisés mais un *tag* aussi fondamental que le nom aurait pu également être pris en compte. Cette proposition construite sur les *tags* spécifiques correspond parfaitement au principe de contribution à OSM. La spécialisation des descriptions est l'un des avantages avancés par les opposants au principe d'un modèle de données strict.

L'évaluation par les qualités intrinsèques des objets n'est pas sans risque d'erreurs. Par exemple, le troisième paramètre de Keßler et De Groot intitulé la « confirmation indirecte » s'appuie sur des hypothèses de comportement de contributions. Considérant un objet, une forte activité contributive autour de cet objet –qui demeure non modifié– le valide. Cette proposition est envisagée par Barron et *al.* (2014) qui ont évalué l'actualité des données en prenant en compte le voisinage de chaque objet. Un deuxième exemple, également proposé par Keßler et De Groot consiste, à l'inverse de la confirmation indirecte, à juger peu fiables les informations issues de corrections attributaires ambiguës (comme une correction du *tag* `amenity=restaurant` vers le *tag* `amenity=cafe`) ou de révisions (comme une création selon un certain *tag* suivi par une autre édition, puis par un retour à l'ancien *tag*). Conscients de la dimension exploratoire de leurs recherches, Keßler et De Groot alertent sur le caractère discutable de ces derniers paramètres. Nous ajouterons qu'une manipulation spécifique de la base de données ne laisse aucune trace dans l'historique de la base de données. L'édition consistant à supprimer un objet et à le remplacer par un nouvel objet qui lui est pourtant quasiment équivalent n'est pas repérable. En outre, ces éditions suppriment tout historique de l'objet qui a été remplacé et certaines de ses caractéristiques intrinsèques en sont altérées. Cette forme de contribution constitue une faille que l'on retrouve dans l'ensemble des recherches évaluant la qualité.

Ce type d'édition, bien que déconseillé dans les préconisations générales, est assez courant. Au sein de Michelin, lors de la mise en place de système de mise à jour des productions, il est ainsi apparu que les numéros de versions n'indiquent pas nécessairement qu'un nouvel objet dans la base de données est effectivement nouveau

de notre point de vue d'utilisateur plus ancien disposant d'une version d'un autre objet quasiment identique. Des systèmes de détection par comparaisons spatiales, de forme et attributaires ont été nécessaires pour distinguer les nouvelles informations des réécritures qui doivent être prises en compte en tant que modifications. Un indicateur entièrement construit sur l'historique des éditions dévalue significativement la création d'un tel objet qui, pourtant, bénéficie souvent des éditions de l'objet qu'il remplace. On remarque parfois ce phénomène pour des tronçons de routes dont la suppression peut être motivée par une géométrie imparfaite. Le contributeur, plutôt que d'ajouter des points et améliorer la courbe du dessin, aura préféré remplacer entièrement le tronçon. Cependant, il reprendra les attributs du tronçon supprimé, s'il les juge exacts. Un dernier exemple des limites des jugements *a priori* positifs de la loi Linus est l'étude menée par Peter Mooney et Padraig Corcoran (2012) sur ce qu'ils appellent les objets « lourdement édités » (souvent modifiés). L'étude ne vise pas directement à l'évaluation des données OSM, mais davantage a exposé le processus contributif, pour des objets rares, régulièrement modifiés, représentant 0,4% de leur jeu de données. Ils concentrent leur analyse sur des objets édités 15 fois ou plus au Royaume-Uni et en Irlande. Si l'analyse de ces objets confirment que le processus itératif réalisé par plusieurs contributeurs est constitutif du développement de la richesse de la base de données OSM, il met également en évidence qu'un objet édité plus de quinze fois l'est en moyenne par moins de cinq contributeurs. En outre, les clés `name` et `highway` sont parfois les motifs de débats que les chercheurs vont jusqu'à qualifier de « guerres d'édition ». Ces guerres peuvent alors mener à des corrections infinies mais sans progression de l'information. Néanmoins, ces corrections incessantes démontrent une vitalité dans la contribution et peuvent être, par exemple, l'indice d'une polémique au sein de la société civile pour la dénomination d'un lieu.

L'évaluation de la qualité des données de la base libre OSM a d'abord été approchée par les méthodes classiques de comparaison à une référence. Les résultats ont rapidement mis en évidence, l'efficacité de ces méthodes mais également le besoin d'une certaine diversification des procédés d'évaluation afin de considérer la pluralité des méthodes de production de données et les différentes dynamiques de contribution locales. Ainsi, Touya et *al.* (2017), proposent-ils de réaliser d'appliquer des méthodes d'évaluation des données successivement. Les chercheurs mobilisent deux méthodes de comparaison à une base de données de référence, une institutionnelle (IGN) et une

folksonomique (Flickr) et ils exploitent également deux méthodes d'évaluation en fonction des caractéristiques intrinsèques des objets, en exploitant d'une part, l'historique de chaque POI analysés et d'autre part, les caractéristiques de leurs objets voisins. Les résultats d'évaluation qualitative confirment les résultats de Haklay et *al.* (2010) en montrant que les modifications surviennent généralement pour améliorer la qualité des objets et réduire leur hétérogénéité dans leurs attributs thématiques. D'in point de vue méthodologiques, Touya et *al.* (2017) concluent qu'une approche holistique est nécessaire pour l'évaluation de la qualité, et considèrent notamment que la recherche de méthodes d'évaluation des données sur OSM doit se poursuivre vers la prise en compte de sources de données davantage diversifiées. Les qualifications folksonomiques que sont les titres, les balises de description des ressources Web sont également envisagées comme des nouvelles sources à envisager. Une autre perspective de recherche (proposée par exemple par Keßler et De Groot (2013)) est la prise en compte de la réputation du contributeur. Existe-t-il des contributeurs experts ? Existe-t-il des contributeurs fiables de manière constante ? Après avoir identifié les caractéristiques intrinsèques des objets, la littérature scientifique s'est intéressée à définir l'expertise et la fiabilité dans le contexte de l'information géographique. Nous définissons le terme d'expertise comme la maîtrise des concepts et des techniques de production de l'information géographique. La fiabilité est une caractéristique évoquant la garantie et la constance du mode de production. De ce fait, un ensemble de données produit par un producteur fiable serait considéré comme d'une qualité constante et s'améliorant (Yang et *al.*, 2016). Un ensemble de données produit par un contributeur expert serait fiable et respecterait des attentes de niveau professionnel.

2.1.3. La crédibilité des contributeurs

Depuis 2010, les chercheurs ont identifié que la densité de données OSM sur un territoire dépend davantage de la densité de contributeurs que de la densité d'informations disponibles à collecter (Girres, Touya, 2010, Neis et *al.*, 2012). D'ailleurs, la liberté totale de renseignement des informations mène rarement à une exhaustivité complète de description d'un territoire. Cette densité de contributeurs devient donc un paramètre de l'évaluation de la qualité des données à part entière. Elle est effectivement associée à une capacité de maintien à jour des données et à la

correction par des contributeurs plus experts des contributions, parfois « ratées », de débutants.

Il convient maintenant de préciser que l'essentiel des données renseignées dans la base de données le sont par une grande minorité des contributeurs (Neis et Zipf, 2012 ; Yang et *al.*, 2016). Ainsi, par exemple, en 2014, 5 % des 85 400 contributeurs allemands, 2,3 % des 32 600 contributeurs français et 3,3 % des contributeurs britanniques avaient édité 90 % des informations présentent dans la base de données OpenStreetMap, depuis 2011 (Yang et al, 2016). À partir de ces résultats, il est acquis que l'essentiel de la base de données est construites par des contributeurs de plus en plus expérimentées, à OSM, d'une part et à la collecte d'informations, d'autre part. Plusieurs études ont tenté d'identifier ces grands contributeurs. Il est apparu que ces rares contributeurs sont également les plus expérimentés. Certains les ont appelés des contributeurs « confirmés » (Neis et Zipf, 2012) et d'autres des « professionnels » (Yang et *al.*, 2016). Ces appellations ne couvrent pas rigoureusement les mêmes caractéristiques.

Le besoin de catégoriser les contributeurs fournit une indication sur leur crédibilité. Ces chercheurs souhaitent mesurer une « réputation » (Keßler et De Groot, 2013) d'un contributeur fiable et compétent. On notera, par exemple, comme caractéristique d'un contributeur crédible, son implication dans l'édition du *wiki* OpenStreetMap (Yang et *al.*, 2016). Ceci ne restera cependant qu'une hypothèse puisque les comptes utilisateurs des deux plateformes ne sont pas partagés. Des résultats semblables étaient déjà connus dans la littérature scientifique. En 2012, Neis et Zipf ont proposé l'appellation de contributeurs confirmés ou « *senior contributors* » (Neis, Zipf, 2012). L'expression souligne l'expertise des contributeurs. Par le terme anglais de *senior* dans l'expression, on identifie la volonté de conférer à ces contributeurs un rôle central dans la contribution. Cette appellation est associée par les deux chercheurs aux contributeurs qui éditent une grande quantité d'objets de la base de données et selon un rythme régulier et constant. Ainsi, dans leur exemple de décembre 2011, les membres du groupe des *senior contributors* étaient les 24 000 contributeurs qui avaient créé plus de mille nœuds depuis leur inscription. L'étude met ensuite en évidence que ces contributeurs sont à l'origine de 89% des *changesets* de la base de données, Alors que ceux-ci ne représentent que 5 % de l'ensemble des personnes inscrites pour contribuer. Enfin, plus de la moitié des contributeurs dits confirmés avait déjà contribué dans un autre pays que leur pays de contribution habituel. Neis et Zipf ont distingué trois autres

groupes. Les 73 000 contributeurs débutants[119] responsables de 14 % de la création de nœuds avaient contribué à au moins dix nœuds. Le plus gros groupe de contributeurs était constitué de contributeurs actifs, dont le nombre de nœuds créé ne dépassait pas dix. Neis et Zipf les désignent « contributeurs irréguliers »[120]. Enfin, ils comptent 312 000 membres qui n'ont jamais été actifs, en ne contribuant jamais tout en disposant d'un compte de contributeur (62%). Yang, Fan et Jing (2016) se sont demandé si ces contributeurs « exceptionnels » ou « confirmés » étaient des professionnels, au sens de personnalités expertes en information géographique, qu'il s'agisse des méthodes, des techniques ou des théories. Différents indicateurs sont utilisés pour identifier ce caractère professionnel : la pratique régulière de la contribution, l'habilité technique et la motivation. Ils permettent d'identifier des preuves solides pour l'expertise (Yang et *al.*, 2016). Il est ainsi observé que des comportements rarement présents chez des amateurs sont repérables chez des contributeurs « majeurs »[121] (Yang et *al.*, 2016). La pratique de la contribution suppose alors l'implication volontaire du contributeur au projet. Elle est mesurée en nombre de jours de contribution : plus ce nombre est important, plus le contributeur a pu développer son habileté à contribuer. Elle est également mesurée par la durée qui sépare l'inscription du contributeur de son dernier jour de contribution qui donne une indication sur sa connaissance du projet et par le nombre de semaines de contribution, qui marque la constance du contributeur. L'habileté technique des contributeurs est aussi mesurée. Elle indique la capacité de la personne à s'intégrer à un domaine spécifique. La compétence à utiliser les outils JOSM, Potlatch et iD est évaluée par les auteurs. Le fait d'utiliser iD n'est pas l'indication qu'il s'agit d'un amateur. Cependant, rares sont les complets amateurs qui utilisent JOSM, peu intuitif et intimidant dans sa grande offre technique. La motivation indique que le contributeur s'implique personnellement dans le projet et s'assure qu'en contribuant, il fournit de meilleures données et les améliore. La contribution, pendant plusieurs jours successifs, est un autre indicateur. Enfin, les auteurs s'intéressent à une plus forte contribution en semaine, signe d'une activité rémunérée. Les résultats montrent clairement que la plupart des contributeurs présentent plusieurs caractéristiques décrites précédemment, indiquant un comportement professionnel. La moitié des contributeurs en compte cinq sur les huit proposés par les auteurs, tandis que

119 « *Junior contributors* »
120 « *Nonrecurring Mappers* »
121 « *Major contributors* »

plus de 75% des contributeurs représentent trois ou quatre comportements. Ce mode d'évaluation présente cependant la limite d'être construit sur la base du comportement des contributeurs, tel qu'il a été identifié et de conduire à des conclusions tirés d'observations partielles.

Ces nouvelles conclusions sont le résultat d'évaluations construites spécifiquement pour répondre à une question qui ne se posait pas aussi régulièrement auparavant puisque la donnée officielle était (et est) considérée comme fiable. Qui est le producteur de l'information que je souhaite utiliser ? Quelle est son expertise ? Un second aspect fondamental attendu du producteur par l'utilisateur est sa fiabilité. Dans quelle mesure est-il éthique ? A-t-il l'intention de me transmettre une information utile, non erronée, non falsifiée ?

2.2. Du projet OpenStreetMap à l'institution OpenStreetMap ?

Le nom OpenStreetMap est celui d'un *projet*. Le terme de projet porte l'idée d'une perspective à réaliser. Le projet OpenStreetMap exprime l'idée de la construction d'une base de données géographique, dont les informations portent sur l'ensemble des espaces occupés, construite et améliorée par un groupe d'internautes connaissant ces espaces. En novembre 2018, la base de données décompressée pèse plus de 1000 gigaoctets (*Planet.osm*, 2018). Il ne s'agit donc plus seulement d'une idée : la réalisation du projet OpenStreetMap, lancée en 2004, s'inscrit maintenant dans la durée. D'ailleurs, la Fondation OpenStreetMap préfère le terme d' « initiative » à celui de « projet » pour décrire OpenStreetMap : « *OpenStreetMap is an initiative to create and provide free geographic data, such as street maps, to anyone* ». Dans cette section, on se demande dans quelle mesure on peut affirmer que l'initiative tend à s'inscrire dans la durée et dans les pratiques. Dans quelle mesure OpenStreetMap peut-elle devenir une source de référence et se transformer en autorité voire en institution dont le mode de fonctionnement pourrait, à un certain moment, devenir traditionnel ?

Pour faire référence, une organisation doit entretenir une relation de confiance avec les récepteurs des informations qu'elle publie. Dans le cas d'informations libres, la relation producteur-utilisateur est troublée comparativement aux modalités traditionnelles de diffusion des données géographiques où la fiabilité d'un producteur est acquise lorsqu'une production crédible apparaît durable et régulière. Le projet OpenStreetMap, dont la production d'informations géographiques est thématiquement diversifiée, est sans équivalent. En tant que producteur d'informations volontaires, OSM n'est pas reconnu *a priori* comme ayant tendance à être crédible par les professionnels de l'information géographique. En effet, ce récent producteur est une communauté internationale numérique non hiérarchisée, dont le mode de fonctionnement est radicalement différent de celui des producteurs traditionnels. La base de données OpenStreetMap est entretenue et améliorée selon des standards souples de création de l'information (Haklay, 2010). Cette souplesse soulève pourtant des interrogations et suscite même de la défiance.

Le mode de production de la base de données OpenStreetMap, par des contributeurs volontaires, est inédit. Son évaluation doit prendre en considération les directions et les tendances prises par les contributeurs afin d'envisager sa crédibilité actuelle et à venir.

Une source de connaissances n'est pas seulement « fiable » par ses résultats et ses acquisitions : elle l'est également par sa capacité à progresser en corrigeant ses erreurs et en se précisant. À travers l'évaluation de la « fiabilité » des données géographiques libres, on cherche à apprécier leur caractère ou non scientifique. Deux concepts proposés par Bachelard (1986) peuvent nous aider dans cette enquête. Le concept de « rectification indéfinie », d'abord, nous permet d'évaluer la capacité du mode de production d'OSM à produire des connaissances les plus rigoureuses, à un moment donné de l'histoire des connaissances et des techniques. Gaston Bachelard explique en effet que « notre connaissance du réel est susceptible d'une rectification indéfinie » (Bachelard, 1986, p. 299). Il voit dans le travail de rectification « la pensée dans son acte, dans son dynamisme profond » (Bachelard, 1986, p.299). Une pensée, ou un domaine de connaissance, qui s'est montré capable de se rectifier, c'est-à-dire de se corriger, de se reprendre et de se transformer, tout en persévérant dans son projet initial, répond le mieux à ce qu'on attend d'une pensée scientifique. L'esprit scientifique cherche à s'approcher de « l'objectivité » du réel, en « limitant le subjectif ». La rectification d'une connaissance est ainsi une limitation et un dépassement de la part de subjectif, au sens de partiel et partial, par un progrès de l'objectivité. Cela exclut une objectivité parfaite. Si l'on suit Bachelard, la force de rectification parcourt la pensée scientifique tout au long de son histoire. Les rectifications qui rythment cette histoire montrent concrètement la dynamique d'un progrès indéfini de la connaissance scientifique. Une source de connaissance qui a su se rectifier, en affinant ses méthodes et en se précisant, peut être considérée comme fiable, étant entendu que la fiabilité est relative à un champ historiquement déterminé de connaissances. Nous nous demanderons, dans cette deuxième section, si ce procédé de rectification est constitutif ou non de la contribution au projet OSM.

Par un second concept, celui de « l'approximation », Bachelard précise les conditions permettant de considérer une connaissance comme fiable, ce que nous entendrons ici par scientifique. Ce concept d'« approximation » ne renvoie pas à un savoir approximatif, vague et sans rigueur, généralement inconscient de ses limites et sans exigence critique. Bachelard parle d'approximation des connaissances scientifiques relatives à un domaine ou un objet d'étude, à l'exemple de l'approximation mathématique (Bachelard, 1986, p.299), c'est-à-dire d'une grandeur que l'on accepte comme suffisamment voisine d'une grandeur connue ou inconnue (que l'on obtient par

estimation, ou troncation). Ainsi, une approximation qualifie une connaissance dont le producteur a conscience du caractère provisoire, mais assez construite et assez utile pour être jugée fiable au moment où on la conseille. L'approximation est ainsi l'application concrète de la rectification indéfinie qui traverse une connaissance scientifique. Elle est « l'objectivation inachevée, mais prudente, féconde, vraiment rationnelle puisqu'elle est à la fois consciente de son insuffisance et de son progrès » (Bachelard, 1986, p.300).

L'approximation est le résultat d'une construction méthodique, fruit d'une rectification de résultats antérieurs, et ouverte à de nouvelles rectifications. En est-il ainsi aujourd'hui des données géographiques libres ? La fiabilisation par une pratique scientifique de la collecte et de l'édition de la base OSM (2.1) est d'abord analysée à travers les processus particuliers soutenant la rectification de la base de données OpenStreetMap. Cette fiabilisation s'opère également par l'acquisition de légitimité (2.2) mise en évidence par l'étude des relations de partenariats scientifiques qui peuvent exister entre les contributeurs à OpenStreetMap et les acteurs traditionnels. Enfin, le mode de production des données OSM soutient la rectification, ouverte à tous les contributeurs. Cependant, c'est cette liberté de rectification qui est le principal facteur de vulnérabilité (2.3) de la base OSM

2.2.1. Fiabilisation par une pratique scientifique de la collecte et de l'édition de la base OSM

Le niveau de fiabilité de la base de données OpenStreetMap peut être appréhendé par l'observation des pratiques guidant sa production. Comme nous l'avons présenté dans la section précédente (II.1), la qualité des données de la base peut être évaluée soit par rapport à une référence, soit du point de vue de leurs propriétés intrinsèques, soit encore par l'évaluation de l'expertise de ses producteurs. Réfutables et comparables à d'autres mesures, les données collectées sont volontairement exposées aux questionnements scientifiques externes. Dans cette sous-section, nous rappelons les principes et les fonctions techniques qui permettent la remise en question de la pertinence des informations retranscrites en données géographiques. Nous nous demandons également dans quelle mesure la collecte d'information pratiquée dans le cadre de l'initiative OpenStreetMap présente un caractère scientifique.

Le fondement théorique du mode de création des informations libres repose sur la dialectique entre la réfutation et la rectification. Comme Olivier Glassey l'explique, dans le cas de Wikipédia, la mise en circulation d'informations libres « entre les membres de la communauté des contributeurs n'est pas univoque mais itérative » (Glassey, 2006, p.239). Il en est de même du mode de production et de diffusion (qui sont simultanés) des données OpenStreetMap. Comme il est expliqué dans la page *Why OpenStreetMap?* « *By the very nature of the wiki-style process there is no guarantee of accuracy of any kind* »[122]. Néanmoins, ce style de création selon le modèle du *wiki* signifie que dans le cas où un contributeur ajouterait, par accident ou par mauvaise intention, une information erronée, d'autres contributeurs seront aptes à corriger l'erreur. Plus précisément, il existe plusieurs degrés séparant une information exacte d'une information erronée. Aux débuts du projet OpenStreetMap l'approximation était reconnue comme inhérente à ce mode de production. L'approximation n'est cependant pas conçue comme dénuée de méthode ni comme irrémédiable. De ce fait, le schéma de la base de données OSM prévoit que chaque nœud, chaque chemin et chaque relation porte un numéro de version. Ce numéro (partant de 1 à la création) de version est incrémenté à chaque nouvelle édition (modification ou suppression) par un contributeur. Depuis octobre 2007, l'historique de chaque objet, organisé selon cette structure incrémentale de versions, est accessible sous la forme du fichier *full_history_planet*, mis à jour de manière hebdomadaire. L'ensemble des versions d'un objet est également accessible au travers de l'interface de requêtes mise en place sur le site *openstreetmap.org*. Par cette traçabilité totale du processus de création des connaissances, le mode de production de la base OpenStreetMap doit être reconnu comme fondamentalement honnête : on estime que le contributeur ne cherche pas à tromper, ni sur les informations qu'il publie ni sur son état d'esprit. Il a conscience des limites de ses connaissances et de ses méthodes. S'il ne prétend pas se soumettre à une éthique rigoureuse, il met en pratique une exigence de transparence et de vérifiabilité, en inventant une technologie d'historisation des informations. Ainsi, le principe de l'édition à OpenStreetMap peut-être résumé par cette phrase de la page *Editing Standards and Conventions* du *wiki* OSM : « *Accuracy. How do you judge and or indicate the accuracy? How accurate is good enough? Is a rough approximation better*

122 Pourquoi OpenStreetMap? « De par la nature même du processus d'écriture en *wiki*, il n'y a aucune garantie de précision qui serait associée à son auteur. Traduction de l'auteure.

than nothing (i.e. inaccurate roads get refined the way wikipedia articles do)?[123] ». L'idée qu'une information approximative est meilleure que l'absence d'information constitue un standard de contribution à OpenStreetMap. Inscrite dans le *wiki* en février 2006, cette phrase est restée inchangée après 130 révisions intermédiaires de la page par plus de 60 contributeurs par rapport à la dernière édition en octobre 2018[124].

Si un fort niveau d'approximation est reconnu comme acceptable et parfois nécessaire, un ensemble de moyens est mis en œuvre pour la **rectification** des données concernées. La page *Quality Assurance*[125] du *wiki* recense plus de 70 outils de contrôle des données. La majorité de ces outils fonctionnent à partir de reconnaissances automatiques d'incohérences. Cependant, les éventuelles incohérences (certaines erreurs sont de faux positifs) doivent être corrigées manuellement[126]. Ces outils peuvent être classés en cinq catégories : déclaration d'erreur ou de manque, surveillance, organisation communautaire, détection d'erreur automatique et documentation.

Le principal outil de **déclaration d'erreurs** est l'outil *Notes. Notes* est intégré au site Web *0penStreetMap.org* et permet à des contributeurs mais aussi à des utilisateurs ne disposant pas d'un compte de contributeur de signaler d'éventuelles erreurs, manques ou imprécisions dans la description d'un lieu.

La catégorie des outils de **surveillance ou monitoring** consiste en un suivi de zones ou de contributeurs. Par exemple, un contributeur, expert d'un territoire, peut utiliser *tyrasd's Latest Changes service* ou *achavi (*construit sur des différentiels entre les fichiers *full_history*) afin de connaître les récents changements concernant une petite zone. Il est également possible d'analyser toutes les contributions d'un contributeur. Il arrive qu'un comportement malveillant puisse ainsi être révélé. C'est la fonction de l'outil *Find Suspicious OpenStreetMap Changesets* du chercheur Pascal Neis. Celui-ci

123 Précision. Comment jugez-vous et / ou indiquez-vous l'exactitude ? Quelle est la précision suffisante ? Une approximation est-elle meilleure que rien (c'est-à-dire que les routes imprécises s'affinent comme le font les articles de Wikipédia) ?

124 Editing Standards and Conventions, consulté en janvier 2019

125 Quality_assurance#Assistant_tools, consulté en janvier 2019

126 La supervision d'un outil automatique par un contributeur est largement encouragée. Néanmoins, des modifications automatisées peuvent ne pas être réalisées directement par un humain. Il peut s'agir de robots, de processus algorithmiques ou de modifications générales via JOSM. Ces modifications massives et incontrôlées une fois débutées peuvent être considérées comme dangereuses bien que permettant d'économiser du temps. On peut lire dans la page du wiki FR:Modifications automatisées (2019) : « Les bots peuvent réaliser des modifications très rapidement et peuvent perturber OpenStreetMap et gêner de nombreux utilisateurs s'ils sont incorrectement conçus ou mal conduits et l'annulation de leur travail peut être compliquée quand ils agissent sur des zones où travaillent activement de nombreux utilisateurs humains. Pour ces raisons, un code de conduite des modifications automatisées a été développé ».

intègre les conclusions de l'article « Towards Automatic Vandalism Detection in OpenStreetMap »[127] (Neis et *al.*, 2012) pour proposer un outil simple d'utilisation et disponible en ligne. On note enfin le référencement de l'outil *OSM Edits Made With MAPS.ME*. Cet outil permet de faciliter l'observation des éditions de la base de données faites à partir de l'application *Maps.me*. Aucun autre éditeur de données n'est contrôlé aussi précisément. Le développement de l'outil d'évaluation et son référencement dans une page aussi centrale du *wiki*, indique une forte méfiance envers l'application *Maps.me* et l'entreprise éponyme la diffusant. Les conditions qui permettent d'expliquer cette méfiance sont détaillées dans le chapitre 3.

Différentes solutions d'**organisation communautaire** sont citées dans la page *Quality Assurance*. Une liste succincte de cartes thématiques comme la carte des monuments, présente une organisation communautaire sur des thématiques précises. Ces regroupements motivés par des intérêts thématiques sont l'occasion de davantage de documentation et d'outils de vérification de la complétude des données. Cette organisation communautaire se traduit également par le développement d'outils de répartition des tâches. L'outil *OSM Tasking Manager* permet de quadriller un territoire afin que les contributeurs, actifs sur celui-ci, puissent indiquer aux autres contributeurs impliqués, l'état de la complétude de leur carré. Ces outils indiquent une recherche de systématisation de la contribution en vue d'une diminution de l'hétérogénéité spatiale.

Les outils automatiques de **détection d'erreurs** sont les plus nombreux de la page *Quality Assurance.* Ces outils sont l'objet de notre discussion sur le processus de rectification comme constitutif du mode de construction des connaissances dans OpenStreetMap. Ils répondent pour certains à des besoins très spécifiques comme la *Tiger Edited Map*, la carte des interdictions de tourner (*turn_restriction*), *OSM Suspects !* et *Unmapped Places of OpenStreetMap* qui concernent des problématiques locales ou thématiques. Ainsi, *OSM Suspects !* Et *Unmapped places* ont tous deux pour objet la mise en évidence d'incohérences typographiques ou attributaires dans la description des implantations de population (ville, village, adresse). Néanmoins, le premier est construit pour le territoire allemand, il prévoit donc des contrôles adaptés aux spécificités allemandes, tandis que le second est un outil dont la logique est rudimentaire mais couvre le monde. D'autres outils comme *KeeRight*, *Osmose* ou

127 L'identification et la rectification du vandalisme sont traitées dans la sous-section II.2.3.

OSMInspector visent à couvrir une large partie des approximations potentielles de contributions sur l'ensemble des zones contribuées. Ces trois outils présentent un niveau de finesse très élevé dans le développement d'algorithmes de reconnaissance d'incohérences ou de potentiels d'améliorations. Les trois outils permettent une redirection vers le site *opensteetmap.org*, pour une correction avec l'éditeur *iD* ou une édition *via* l'éditeur *JOSM*. *Osmose* et *OSM Inspector* permettent également la déclaration de faux positifs. Parmi leurs nombreuses fonctions, on relève les fonctions d'erreurs topologiques (superposition de polygones de même type, direction des tronçons incohérente, routes non connectées), d'erreurs attributaires (*tags* inconnus, valeur incompatible avec la clé, classification de la hiérarchie des villes incohérente, utilisation de *tags* dépréciés) et d'avertissement de possibles incohérences (forte densité de nœuds en un même endroit, langues non reconnues). Osmose propose l'ensemble de ces aides à la correction (fig. 21).

L'outil Osmose présente pour nous un intérêt encore plus vif car il intègre des sources de données tierces à partir desquelles sont proposées des rectifications. Ces données tierces sont le plus souvent gouvernementales et elles sont toutes diffusées sous une licence compatible à l'ODbL, leurs présences marquent des premiers partenariats entre institution et OSM. En outre, l'outil est conçu pour développer des contrôles adaptés aux spécificités territoriales. En France, on note l'usage des référentiels des ministères, ainsi que ceux de l'INSEE, comme moyen de confrontation des informations. Osmose recourt à Twitter pour mettre en avant la capacité d'OSM à s'ouvrir à des problèmes posés en dehors du territoire français comme les limitations de vitesse en Allemagne. Par ailleurs, Osmose s'appuie sur les algorithmes de reconnaissances automatiques de signaux présents dans les photographies de la base de données *Mapillary* pour proposer des rectifications aux données OSM. Enfin, dans la page *Quality Assurance*, on note la citation de deux outils diffusés par de grandes entreprises : *Improve OSM* de Telenav et *OSM-Sidewalker* de Mapbox[128].

128 Le rôle des entreprises dans le développement de la base de données OpenStreetMap est traité dans la sous-section suivante, dans le chapitre 3 et le chapitre 4.

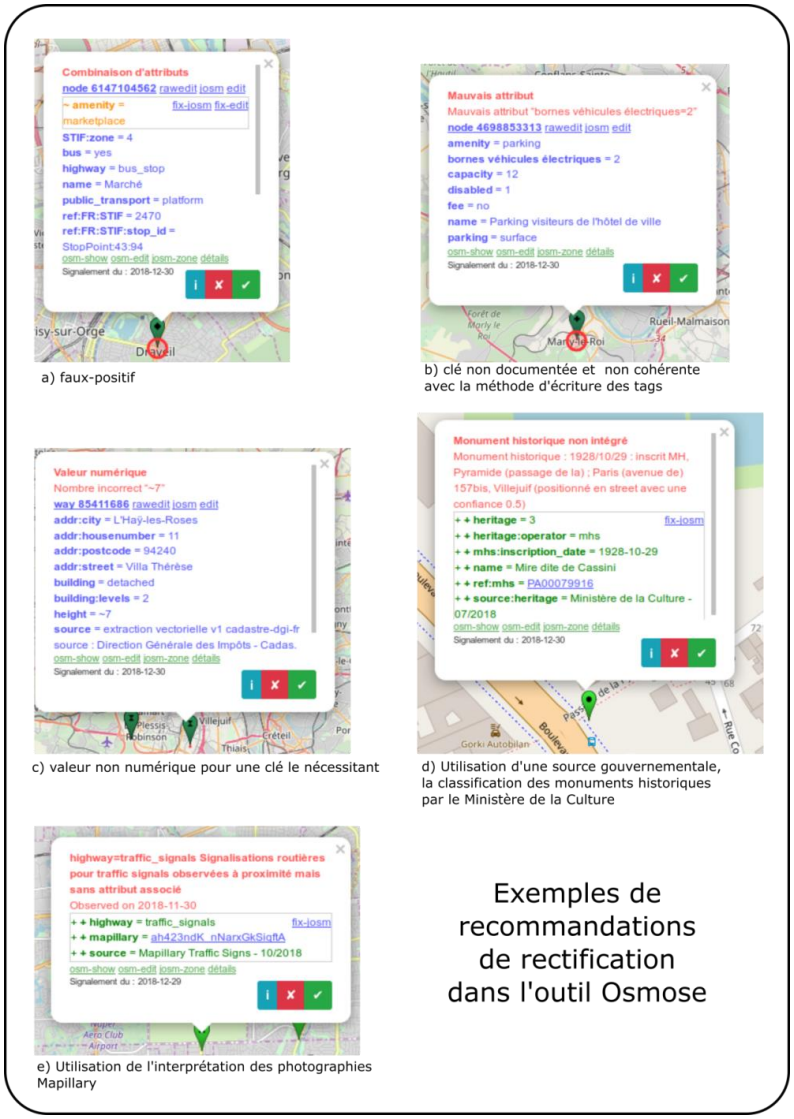

Figure 21 Diversité des résultats des analyses automatiques offerts dans Osmose (Hayat, 2019)

La cinquième, et dernière, catégorie d'outils développés pour la rectification et la diminution de l'approximation relève de la **documentation**. La documentation de la contribution à OpenStreetMap présente deux aspects. D'une part, une partie de cette documentation détaille des méthodes rigoureuses de collecte de l'information. D'autre part elle est, par le moyen du *wiki*, un moyen de discussion et d'amélioration. La rigueur de la collecte recommandée au travers de la documentation OSM présente un caractère scientifique. La première page qui nous intéresse est intitulée *Pick your mapping technique*. Il s'agit en fait de la troisième page du guide du débutant, très visible à l'ouverture du *wiki* OpenStreetMap. L'exigence scientifique est repérable au travers de l'exposition des techniques possibles et de leurs bons usages, à travers l'énumération

des sources disponibles et du respect des licences ainsi qu'à travers de l'explicitation de la méthode générale qui repose sur une approximation contrôlée et transparente. Deux types de techniques y sont conseillés : la contribution à partir de sources documentaires (images satellites, photographies et données tierces) et la contribution par relevés de terrain. Les connaissances personnelles et les souvenirs peuvent être utilisés, s'ils sont récents. Les suppositions ne sont pas encouragées et l'approximation bien qu'acceptée doit être transparente[129]. La contribution réalisée uniquement à partir de sources documentaires est appelée « *armchair mapping* » c'est-à-dire cartographie dans un fauteuil. Cette forme de création d'informations est reconnue comme utile. Elle constitue néanmoins le procédé le plus exposé à l'approximation. Le nombre de voies, le revêtement d'une route, les toponymes ou les équipements sont autant d'informations rarement lisibles au travers de sources disponibles librement. Celles-ci sont cependant considérées comme des sources secondaires confirmant des relevés[130]. En citant les sources que sont les imageries de *Bing*, les photos des bases *Mapillary, OpenStreetCam, Bing Streetside, Pic4carto* et les traces GPS enregistrées par les contributeurs, un socle commun de ressources documentaires est déclaré dans cette page. Cette page introductive débute par l'explication de la technique de l'*armchair mapping* sans oublier d'encourager le contributeur débutant à la pratique de la sortie de terrain. Néanmoins, la difficulté à s'acclimater à la contribution est prise en compte. Évoquée comme la suite logique de l'*armchair mapping*[131], la difficulté de la sortie de terrain, n'est pas masquée. Il est précisé qu'il n'est pas requis de maîtriser toutes les techniques présentées pour contribuer et celles-ci sont exposées selon une gradation vers des techniques de plus en plus expertes au fil du texte. La présentation commence par l'application permettant la contribution via des formulaires (*StreetComplete*) jusqu'à l'import de traces GPS dans JOSM. Cette progression évite d'exclure un nouveau contributeur non expert et évite aussi l'emploi inadapté d'outils complexes avant acclimatation au procédé de contribution. Enfin, sont décrites des méthodes scientifiques classiques de collecte, comme la prise de notes en parallèle

129 À ce propos, on peut lire dans la page du wiki Pick your mapping technique : « *avoid guessing. Mark estimations with note=* or fixme=*. If unsure, use special general tags like highway=road or building=yes or shop=yes instead of their specific variants* ».

130 « *Once you are comfortable with OSM, consider mixing [Aerial imagery] with other techniques* », *Pick_your_mapping_technique* (2018).

131 « *You are highly advised to incorporate outdoor mapping once you are comfortable with OSM* », *Pick_your_mapping_technique* (2018).

d'enregistrements de points GPS. La technologie GPS n'est pas présentée comme étant infaillible. Ainsi, des mesures répétées pour un même tracé sont encouragées afin de s'assurer de leur régularité.

Dans le guide du débutant, la page *Pick your mapping technique* est précédée de la page *OSM data explained*[132]. Cette dernière page résume les conventions décidées par les contributeurs votants qui résultent du processus d'itération que nous avons expliqué au début de cette sous-section[133]. La tendance à la rectification s'observe également au travers de la précision progressive de la *folksonomie* et de la documentation. Mocnik et *al.* (2017) montrent que, le plus souvent[134], un nouveau *tag* apparaît d'abord comme contribué à la base de données avant d'être décrit dans le *wiki*. Les chercheurs ont évalué que malgré ce décalage, paradoxal du point de vue d'une production classique, les pages des *tags* étudiées (« *the relevant tags* ») « *seems to be close to completion* » dans le *wiki*. Afin de considérer le niveau de documentation des *tags*, ils choisissent donc un seuil à partir duquel la définition est nécessaire dans le *wiki*. Le seuil de 1000 contributions dans la base est choisi. Celui-ci correspond à l'observation selon laquelle un *tag* comme `name=New-York-City` atteint un niveau de spécialisation bien supérieur au niveau attendu dans une documentation. Présentant des phases d'évolution de la folksonomie et en parallèle de sa documentation, Mocnik et *al.* présentent trois phases. La première phase est celle du début de la folksonomie, durant laquelle la documentation est très limitée. Durant la deuxième phase, la documentation s'accroit fortement, folksonomie et documentation tendent à correspondre. Les chercheurs placent le projet OSM dans la troisième phase durant laquelle la folksonomie se précise et ce raffinement de granularité se retrouve dans la documentation. À ce stade, Mocnik et *al.* estiment qu'il est donc possible d'étudier la folksonomie d'OSM à travers son *wiki*. Les résultats des chercheurs paraissent corroborer notre hypothèse selon laquelle **le processus de production de connaissances dans le cadre de l'initiative OpenStreetMap est basé sur un fort processus de rectification.**

132 Consulté en janvier 2019

133 « *Tagging rules and conventions change with time. These conventions are decided by the community through common usage, voting, and discussion attalk:tagging. Here is a list of important tag keys that you should be aware of. Most editors have helpful presets that will help you get started easily, without knowing tags in depth. But knowing the basics certainly helps* », Beginners Guide 1.3., consulté en janvier 2019.

134 Comme les auteurs l'expliquent, il n'existe pas de stricte limite selon laquelle on pourrait affirmer qu'un *tag* est utilisé et devrait donc être documenté. D'après leurs visualisations, la majorité des *tags* ne sont pas documentés après 100 utilisations.

Deux autres paramètres d'une production scientifique peuvent intervenir pour apprécier OpenStreetMap : la **reproductibilité** des résultats et leur **vérifiabilité**. Dans le contexte de la production d'une base de données, la reproductibilité des résultats signifie que les moyens par lesquels la collecte de l'information a été réalisée sont explicités et qu'une nouvelle collecte mènerait à des résultats identiques. Différentes techniques de collecte sont expliquées dans les pages du *wiki* et une large part des *tags* est explicitée afin d'aider les contributeurs à choisir les plus adaptés aux objets qu'ils décrivent. Cependant, la reproductibilité de la collecte ne peut être assurée. L'approximation et la subjectivité (au sens du choix du niveau de détail attributaire et géométrique) des contributeurs est inhérente à la production libre. C'est par le processus de rectification que le caractère scientifique est assuré. Enfin, compte tenu du caractère topographique de la base de données, cantonnée aux espaces publics et semi-publics, la vérifiabilité des données est assurée en théorie. Comme Keßler et de Groot (2013) l'ont démontré, une enquête de terrain permet de vérifier les données de la base OSM. Néanmoins, il est impossible de vérifier manuellement l'ensemble des données de la base OSM. C'est pourtant là tout l'enjeu des *fake news.*

2.2.2. Fiabilisation par l'acquisition d'une légitimité

La production de la base de données OpenStreetMap présente certaines caractéristiques d'une production scientifique (la production de la connaissance est réalisée selon le processus de rectification, les résultats de collecte doivent pouvoir être reproduit par un autre contributeur et chaque information doit pour être vérifiée). De ce fait, on peut se demander si l'initiative OpenStreetMap appartient au champ scientifique de l'information géographique. En 1975, Pierre Bourdieu définit le champ scientifique comme « un champ social comme un autre, avec ses rapports de forces et ses monopoles, ses luttes et ses stratégies, ses intérêts et ses profits, mais où ces invariants revêtent des formes spécifiques » (Bourdieu, 1975). Il soutient que le fonctionnement d'une science repose sur des luttes pour la légitimité et pour le « monopole de l'autorité scientifique »[135], associé à cette légitimité. En considérant les rapports entre les acteurs

135 « le champ scientifique comme système des relations objectives entre les positions acquises (par les luttes antérieures) est le lieu (c'est-à-dire l'espace de jeu) d'une lutte de concurrence qui a pour enjeu spécifique le monopole de l'autorité scientifique inséparablement définie comme capacité technique et comme pouvoir social, ou si l'on préfère, le monopole de la compétence scientifique, entendue au sens de capacité de parler et d'agir

(dans notre cas les institutions et OpenStreetMap) d'un champ scientifique comme des rapports de classes, la « lutte scientifique » s'exerce entre des dominants et des dominés. Les dominants imposent « la définition de la science selon laquelle la réalisation la plus accomplie de la science consiste à avoir, être et faire, ce qu'ils ont, sont ou font ». Toujours selon Bourdieu, l'acteur dominant tend à imposer une définition de la science qui sera « la plus conforme à ses intérêts spécifiques, c'est-à-dire la mieux faite pour lui permettre d'occuper en toute légitimité la position dominante en assurant la position la plus haute dans la hiérarchie des valeurs scientifiques aux capacité scientifiques dont il est le détenteur à titre personnel ou institutionnel ». Si l'on suit Bourdieu, la lutte scientifique sert les intérêts des acteurs. Bourdieu rompt avec l'idée selon laquelle la science « ne connaîtrait pas d'autres lois que celle de la concurrence pure et parfaite des idées » et considère que le fonctionnement d'un champ scientifique « suppose une forme spécifique d'intérêts ». Ces intérêts seraient masqués par leur prétendu « désintéressement exigés par d'autres champs » et, notamment un désintéressement économique. Se fondant sur l'idée selon laquelle une activité scientifique vise « l'acquisition de l'autorité scientifique (prestige, reconnaissance, célébrité, etc.) » Bourdieu affirme que « dans le champ scientifique [...], il n'existe pas d'instance à légitimer les instances de légitimité ; [...] personne n'est *bon* juge parce qu'il n'est pas de juge qui ne soit juge et partie ».

La théorie du système du champ scientifique de Bourdieu expose la problématique du monopole dans une science. En cas de monopole, le renouvellement de la science par changement de paradigme semble inaccessible. Néanmoins, Bourdieu n'exclut pas la possibilité de l'émergence d'une concurrence. La légitimité est revendiquée « par un groupe » qui entre en rapport de force avec le groupe dominant. Afin de répondre à notre question de la fiabilité d'OSM, on s'intéresse ici à la possibilité pour les contributeurs à OSM de construire leur légitimité. On se demande si OpenStreetMap représente un acteur concurrentiel pour les autorités scientifiques du champ de l'information géographique. Dans la mesure où OSM n'est pas un acteur précisément identifiable et que sa propre définition politique est encore à l'état de projet, existe-t-il nécessairement une lutte entre acteurs pour le monopole scientifique mettant en jeu d'une part les méthodes de production de l'information et d'autre part le rapport à

légitimement (c'est-à-dire de manière autorisé et avec autorité) en matière de science, qui est socialement reconnue à un agent déterminé » (Bourdieu, 1975 pp. 91-92).

l'expertise en information géographique ? OSM peut-elle apparaître comme une autorité reconnue, alors qu'elle ne prétend pas en être une et qu'elle ne s'est en conséquence pas impliquée dans des luttes pour s'imposer comme une autorité ?

Initialement perçue comme produite par des amateurs, sans respect de normes et diffusée sous une licence « virale », la base de données OpenStreetMap est devenue un objet d'étude scientifique lorsqu'en 2007, Michael Goodchild donne un nom à son mode de production, le VGI (*Volunteered geographic information*). Les premières formes de reconnaissance sont alors repérables lorsque des chercheurs en informations géographiques comparent OSM à des bases de données de référence (c'est-à-dire produites par des *autorités scientifiques*). Différentes autres formes de reconnaissance de l'utilisabilité de la base de données OSM dans des contextes comparables aux bases de références sont également à relever. Ainsi, l'usage progressif de fonds de carte en ligne issus de la base de données OpenStreetMap n'est pas mesurable mais se constate clairement. On retient, par exemple, l'usage des services de fonds de carte de l'entreprise Mapbox[136] par des entreprises de services numériques comme SnapChat, The Lonely Planet, The Weather Channel ou National Geographic.

On peut également considérer les différentes formes de partenariats entre institutions et OpenStreetMap comme le signe de l'intégration du projet dans la sphère des autorités du domaine et sa reconnaissance par cette même sphère. La forme la plus élémentaire du partenariat est celui des *imports* de données d'une institution dans la base de données OSM. La page *Import/Catalogue* du *wiki* OpenStreetMap recense ainsi 289 imports de « grandes échelles » de données tierces. Une certaine exhaustivité de ce recensement est assurée par l'exigence de se déclarer dans la page *Import/Guidelines*[137]. Différentes formes de partenariats sont à distinguer. Le partage de données avec OSM se produit :

– par la diffusion des données dans le domaine public,

– par l'usage d'une licence compatible voire même par le choix de la licence ODbL,

– par une entente spécifique entre le producteur et la Fondation OSM ou l'association OSM locale.

[136] Nous détaillerons dans le chapitre 3 les usages commerciaux des données OpenStreetMap.
[137] « *You must register your permissions and project by adding a line to the table at* Import/Catalogue ». *Import/Guidelines*. (2018)

Parmi les premiers imports massifs de données publiques dans la base OSM, on retient l'exemple de l'import, en 2007, du réseau routier issu de la base de données *TIGER*, datant de 2005, publié dans le domaine public par l'*US census bureau TIGER* data. Le réseau routier d'origine gouvernemental couvrant le territoire états-unien a été intégré à la base de données OSM. En moins de six mois, il apparaît que l'ensemble du pays est couvert par un réseau dense alors que ce réseau était quasiment inexistant dans la base de données OSM (fig. 22).

Figure 22 L'import du réseau TIGER *en septembre 2007 à gauche et le réseau à la fin de l'import en janvier 2008*

S'il s'agit d'un partenariat encore modeste, la publication dans le domaine public est un choix politique fondé sur la reconnaissance de la capacité des citoyens à se saisir de ces informations. Cependant, la qualité de la géométrie des objets et la qualité attributaire sont variables. Par exemple, la classification du réseau *TIGER* n'était pas compatible avec la classification du réseau OSM. De ce fait une partie importante du réseau urbain a été classé, à l'import, selon le *tag* `highway=residential`. Selon la classification propre à OpenStreetMap, le *tag* `highway=residential` peut être adapté mais plusieurs autres cas exigent l'emploi des *tags* `highway=tertiary` ou `highway=service`. Ces trois *tags* représentent des objets assez éloignés dans la hiérarchie d'un réseau. Afin de soutenir les contributeurs pour qu'ils participent à la correction du réseau *TIGER* dans OSM, une page largement documentée du *wiki*[138] OSM intitulée « TIGER_fixup » (correction de *TIGER*) est mise en ligne le 20 septembre 2007. Elle est encore active en janvier 2018. En 2013, Dennis Zielstra, Hartwig H. Hochmair et Pascal Neis ont étudié les effets de l'import de cette base de données sur la progression de la complétude du réseau routier états-unien. Ils notent

[138] TIGER fixup. (2018), consulté en janvier 2019.

que les communautés de contributeurs états-uniennes ne sont que modérément actives à l'édition pour la correction et la mise à jour des données importées. Les auteurs encouragent la communauté à développer des outils exploitant les nouvelles publications de *TIGER* au service de mises à jour manuelles et non d'imports massifs. En 2014, ces résultats sont confirmés par Grochenig et *al*.. Ceux-ci montrent de forts écarts de rythme de contribution localement entre des zones ayant bénéficié d'imports massifs de bâtiments ou de rues où la contribution par relevés de terrain est faible (peu dynamique en 2013) et les zones dynamiques, dont l'ensemble de la donnée provient de contributions manuelles et rarement d'imports de données. La communauté des contributeurs, également consciente des dérives d'un tel import, s'appuie sur cet exemple pour développer des règles en faveur d'une interaction plus forte avec les contributeurs. D'ailleurs, l'exemple de l'import de la base TIGER est cité à six reprises dans la page *Import/Guidelines* pour illustrer les règles suivantes : « *Use the right tags* », « *Don't put data on top of data* », « *Keep server resources in mind* ». La tendance à publier des données selon des licences libres s'est développée de façon rapide et ces publications ont été l'occasion d'imports massifs de données, selon des procédures de contribution strictement définies. Nous retenons, en 2009, les imports des limites administratives de l'Espagne et de l'Italie, disponibles sous des licences exigeant l'attribution des données aux producteurs que sont *l'Instituto Geográfico Nacional* en Espagne et *l'Istituto nazionale di statistica* en Italie ; en 2013, suite à la création d'un projet de publication de donnés ouvertes relatif aux numéros de rue, sous licence CC BY 3.0 par la ville de Florence, la communauté en réalise l'import; en 2018, la publication en ODbL des bâtiments par le *Centre for GIS and National Bureau of Statistics* des Seychelles a permis l'import de 32000 *buildings*[139].

Enfin, la dernière forme de partenariat, est selon nous, une réelle forme de reconnaissance de l'utilité et de l'utilisabilité de la base de données OSM. Il s'agit de l'entente écrite et juridique. Par le moyen des licences, les institutions publiques ont donné accès à une partie, ou à l'ensemble de leurs bases de données. De ce fait, la liste des imports effectués dans OSM présente dans ses caractéristiques les conditions de licence qui ont permis chacun des imports. Les conditions intitulées « *Explicit*

[139] « *Dear OSM-Communit, The import of about 32,000 buildings is completed. Best regards, Michael Michael Wagner Geospatial Software Developer and Consultant all spatial - Geospatial Software and Services* ». Message envoyé sur la liste de discussions, relatives aux imports de données dans OpenStreetMap. Consulté en ligne en janvier 2019 : https://lists.openstreetmap.org/pipermail/imports/2018-August/005658.html

permission to use without any restriction » ou « *Licence Agreement* » correspondent à ces situations d'entente solides entre partenaires. Parmi ces différents partenariats, on peut citer la mise à disposition, en 2013, au profit d'OSM[140], des bâtiments référencés sur la ville de Dacca au Bangladesh par *The Center for Environmental and Geographic Information Services* (CEGIS). Des projets de partenariats de plus grande ampleur peuvent être également signalés, comme celui de la mise à disposition du Cadastre, en France, par la Direction générale des finances publiques en janvier 2009. Ainsi, comme il est expliqué sur la page dédiée aux droits d'utilisation, *WikiProject Cadastre Français/Conditions d'utilisation* (2018), en 2009, « la DGFiP autorise les contributeurs OSM à accéder directement au serveur WMS dont l'ouverture au public était de toute façon programmée pour le milieu de l'année 2009. Cela a toutefois permis le développement et l'utilisation anticipée du *plugin* cadastre-fr pour JOSM ». En début de page, il est précisé que désormais, « les données du cadastre sont disponibles sous Licence Ouverte (*Open Licenc*e) sur le site d'Etalab (portail o*pen data* français) ». Elles sont donc réutilisables à condition d'en citer la source et le millésime.

Une conséquence de ce partenariat est l'import semi-automatique des bâtiments vectorisés de la base cadastrale. D'une plus grande ampleur encore, l'agence du gouvernement fédéral canadien responsable des statistiques (*Statistics Canada*) a mis en place un projet exploratoire qu'elle présente sur son site[141]. Présenté depuis 2016 dans le *wiki*, ce projet a débuté à Ottawa. Cette page est principalement écrite par le contributeur canadien *James2432* qui explique l'intérêt que porte l'agence des statistiques pour OSM. Ceux-ci ont identifié OSM comme un moyen d'acquérir des données thématiques à l'échelle des rues de la ville. Cette description est cohérente avec celle présentée par l'agence elle-même qui explique : « *This exploratory initiative aims at enhancing the use and harmonization of open building data from government sources for the purpose of contributing to the creation of a complete, comprehensive and open database of buildings in Canada.[...] This initiative originates from insights taken from the Statistics Canada pilot project on data crowdsourcing, which used OpenStreetMap as a platform for integrating data on building footprints. In addition to the possible benefits of crowdsourcing, that project highlighted the potential of integrating open*

140 « The data is provided by CEGIS, who has extended explicit permission to import it into OpenStreetMap » Dhaka Building Import (2019), consulté en janvier 2019.
141 *Open Building Data: an exploratory initiative*: https://www.statcan.gc.ca/eng/open-building-data/index, consulté en mai 2019.

data from municipal, regional, and provincial governments to meet the needs of official statistics»[142]. Ainsi, afin de soutenir la contribution et dans un souci d'exhaustivité, mais également d'interopérabilité, la ville d'Ottawa, en discussion avec les contributeurs locaux, s'est assurée d'une compatibilité entre les licences de ses données et celle des données OSM. Il apparaît que 259 336 `buildings` portent le *tag* `source=City_of_Ottawa`. Ces bâtiments ainsi *taggés* représentent 92.67% des objets de la base qui portent ce *tag*[143]. Ce dernier exemple se situe à la frontière entre un partenariat passif de diffusion de données libres et un partenariat actif en faveur des données libres.

Depuis le début du projet OSM, des ententes spécifiques entre des producteurs publics et privés et la Fondation OSM ou l'association OSM locale ont été mises en place. On peut ainsi citer le contrat entre Bing et la Fondation OSM octroyant le droit aux contributeurs à OSM de dessiner de nouveaux objets dans la base OSM à partir de ses images. L'entreprise Mapbox procède au même partenariat exclusif avec OSM, offrant l'accès aux contributeurs à ses images satellites. En France, le 20 mai 2016, l'IGN signe une convention avec OSM France autorisant les contributeurs OpenStreetMap à utiliser les images aériennes de la BD Ortho pour contribuer à OpenStreetMap. En janvier 2019, près de 200 000 objets de la base de données OSM portent un *tag* indiquant que la source du dessin de cet objet est la BD Ortho de l'IGN. Plus largement, on compte dans la page du *wiki* Aerial imagery, plus de 30 partenariats de ce type. Ce recensement n'est pas exhaustif (la BD Ortho n'est d'ailleurs pas signalée) et les dates des images ne sont pas renseignées. Chacun de ces exemples de partenariats a pour visée une meilleure production de données et l'ouverture des données publiques.

On peut lire dans ces échanges une forme de conciliation avec les autorités scientifiques. Des partenariats plus aboutis ont également été envisagés. À un niveau institutionnel, on citera, en France, le cas de la Base Adresse Nationale (BAN) constituée, suite à la signature d'une convention le 15 avril 2015, par la collaboration

[142] Ce projet exploratoire vise à améliorer l'utilisation et l'harmonisation des données ouvertes relatives aux bâtiments, provenant de sources gouvernementales, dans le but de contribuer à la création d'une base de données complète et ouverte sur les bâtiments au Canada. [...] Ce projet découle des connaissances tirées du projet pilote de *crowdsourcing* du bureau national de statistique, qui utilisait OpenStreetMap comme plate-forme d'intégration de données sur l'emprise des bâtiments. En plus des avantages possibles du *crowdsourcing*, ce projet a mis en évidence le potentiel du processus d'ouverture des données provenant des administrations municipales, régionales et provinciales pour répondre aux besoins en statistiques officielles. Traduction de l'auteure.

[143] https://taginfo.openstreetmap.org/tags/source=City%20of%20Ottawa#combinations, consulté en mai 2019.

entre Etalab, La Poste, l'IGN, la DGFiP et OpenStreetMap France[144]. La BAN est le résultat du rapprochement des données et traitements adresses de ses partenaires. Sa constitution répond aux objectifs d'interopérabilité entre les bases de données adresses déjà existantes. En constituant une base nationale unique d'adresses, maintenue par des organismes publics, les partenaires visent à couvrir le territoire et à garantir des données actuelles. Le projet de la BAN partage en fait l'objectif de celui du projet de la Base Nationale Ouverte (BANO) initiée en octobre 2013 par OSM France. Le projet BANO est décrit dans sa page dédiée du *wiki*, comme une initiative d'OpenStreetMap France. « Elle a pour objet la constitution d'une base libre la plus complète possible de points d'adresse à l'échelle de la France »[145]. S'intégrant dans l'écosystème d'OpenStreetMap, la BANO est uniquement publiée sous la licence ODbL. Cependant, alors que la BAN est publiée, actuellement sans licence, la plaçant sous les règles de droit commun du Code des Relations entre le Public et l'Administration (CRPA), il est impossible d'y importer des données sous licence ODbL. En effet, dans le cas de la constitution d'une base de données composite en ajoutant des données ODbL, la BAN devrait également être publiée sous cette licence *copyleft*. Face à ce problème, d'une part, la BAN est également publiée sous une licence ODbL, permettant le croisement avec des données sous ODbL et d'autre part, (BANO) reste d'actualité. On peut considérer que ce partenariat institutionnel n'est que partiellement effectif. Cependant, deux éléments nous indiquent une prise en considération du rôle d'OSM dans la production d'informations géographiques sur le territoire français. D'abord, OpenStreetMap est un membre signataire de la convention qui a ouvert la création de la BAN. Ensuite, la publication en ODbL de la BAN indique une volonté d'interopérabilité avec la base OSM mais également avec d'autres bases de données qui auraient été placées sous la même licence.

D'après ces exemples, on peut considérer qu'OSM est en relation avec des autorités du domaine de l'information géographique. Néanmoins, ce rapport qu'entretient OSM avec les autorités de son domaine n'est pas rigoureusement celui décrit par Pierre Bourdieu. En effet, l'entité floue qu'est OSM ne prétend aujourd'hui à aucune

144 « La Base Adresse Nationale est une base de données qui a pour but de référencer l'intégralité des adresses du territoire français. Elle contient la position géographique de plus de 20 millions d'adresses. Elle est constituée par la collaboration entre Etalab, La Poste, l'IGN, la DGFiP et OpenStreetMap France. », (Base Adresse Nationale BAN, data.gouv.fr, Consulté en mai 2019).

145 FR:WikiProject France/WikiProject Base Adresses Nationale Ouverte (BANO). (2019), consultée en janvier 2019.

légitimité politique ou de service public. Le groupe de contributeurs, les associations locales et la Fondation OSM expriment leur légitimité comme une réponse à des besoins ignorés par ces services publics. C'est pourquoi la liberté de description est si fondamentale dans le projet. Si un contributeur ou un groupe de contributeur estiment qu'un ensemble d'informations géographiques devraient être disponible il leur est possible de le créer. Ce positionnement, en décalage avec les autorités actuelles, élimine partiellement l'entrée dans un rapport de force. En outre, le rapport de forces décrit par Bourdieu s'exprime en partie par une censure des inventions nouvelles en quête de légitimité. Dans notre contexte, cette censure n'est pas possible. Le contexte de publication qu'est Internet exclut la censure. En effet, OSM ne participe pas de l'évaluation par les pairs. Ce processus étant évité, la censure n'est pas un facteur potentiel d'expression de conflit.

Le rapport de force n'est pas annulé. On retient deux exemples révélateurs. Le premier est celui qu'on peut appeler le « récit fondateur d'OpenStreetMap ». Ce récit est celui de l'insatisfaction ressentie par Steve Coast, en 2004 lorsqu'il demande à l'*Ordnance Survey*, l'agence cartographique du Royaume-Uni, l'accès gratuit à un jeu de données géographiques. Steve Chilton, *chairman of the Society of Cartographers*, le raconte en 2009 dans les *Proceedings* de la 24e conférence internationale de cartographie. « *OpenStreetMap (OSM) was started in the UK by Steve Coast in August 2004, because of his frustration with the tight copyright that pertains to the Ordnance Survey (UK's National Mapping Agency) maps and data. In his view this was restricting him from producing a local map, and at a macro level stifling innovation by not making the map data available to users without excessive royalty payments. So with a consumer grade GPS he started collecting tracks around his local area of central London, and writing some reasonably unsophisticated software to display this data. […]. Shortly after presenting his idea at the EuroFOO conference he started getting others interested and contributing data in like mode* »[146] (Chilton, 2009). Ce récit est repris dans différents articles scientifiques comme élément de contexte explicatif des motivations du projet

146 OpenStreetMap (OSM) a été lancé au Royaume-Uni par Steve Coast en août 2004, en raison de sa frustration vis-à-vis des droits d'auteur qui restreignent l'accès aux données géographiques de l'Ordnance Survey (agence de cartographie nationale du Royaume-Uni). Selon lui, cette restriction l'empêchait de produire une carte locale. En ne mettant pas les données cartographiques à la disposition des utilisateurs sans paiement excessif de redevances l'Ordnance Survey étouffait l'innovation. Avec un GPS grand public, il a commencé à collecter des pistes cyclables autour de son quartier du centre de Londres et à écrire un logiciel simple pour afficher ces données. Peu de temps après avoir présenté son idée à la conférence EuroFOO, il a commencé à intéresser les autres et à fournir des données de la même manière.

(Chilton, 2009; Glasze et Perkins, 2015). On retrouve aussi ces motivations à la page d'accueil du *wiki* OSM : « *Welcome to OpenStreetMap, the project that creates and distributes geographic data for the world. We started it because most maps you think of as free actually have legal or technical restrictions on their use, holding back people from using them in creative, productive, or unexpected ways*[147] » (Main Page, 2014). Ce récit fondateur est également évoqué en ouverture de la page *History of OpenStreetMap* du *wiki*; « *Steve Coast founded OpenStreetMap in 2004, initially focusing on mapping the United Kingdom. In the UK and elsewhere, government-run and tax-funded projects like the Ordnance Survey created huge data sets, but failed to freely and widely distribute them* »[148]. Le second exemple d'un rapport de force entretenu entre OSM et les institutions, que nous retenons est l'échange qui a eu lieu, en 2013, dans la Gazette des Communes entre Gaël Musquet alors président d'OpenStreetMap France et l'IGN représenté par Pascal Berteaud, directeur général de l'IGN. L'objet du conflit est le rapport Trojette relatif aux redevances sur les données publiques. Dans le cadre du processus d'ouverture des données publiques, le rapport de la Cour des comptes pose la question de la légitimité des exceptions au principe de gratuité. Le rapport rappelle le principe de l'*open data* selon lequel la publication de données publiques soutient la démocratie et l'innovation. Face à ce projet d'ouverture des données, des services publics ont mis en place ou maintenu des redevances en échange de leurs données pour des motifs de « besoin budgétaire, situation de monopole, vision patrimoniale des informations et de la ressource qui en est tirée, volonté de limiter la demande ou de protéger un écosystème existant, une vingtaine de services publics administratifs ont institué ou maintenu des redevances de réutilisation ». Le rapport Trojette pose la question de la légitimité de ces redevances et prône « la transition vers de nouveaux modèles économiques » encourageant l'État à assurer cette transition[149]. Suite à la publication du rapport, le 5 novembre 2013, la Gazette des Commune publie le même jour les réactions de sept personnes ayant été

147 Bienvenue dans le projet OpenStreetMap, le projet qui crée et distribue des données géographiques pour le monde. Nous l'avons créée parce que la plupart des cartes que vous considérez comme gratuites ont en réalité des restrictions légales ou techniques, empêchant les utilisateurs de les utiliser de manière créative, productive ou inattendue.

148 Steve Coast a fondé OpenStreetMap en 2004, se concentrant initialement sur la cartographie du Royaume-Uni. Au Royaume-Uni et ailleurs, des projets gérés par le gouvernement et financés par les impôts, tels que l'Ordnance Survey, ont créé d'énormes ensembles de données, mais n'ont pas réussi à les distribuer librement et largement.

149 « Les coûts de production et de collecte des informations publiques devraient, en toute rigueur, peser exclusivement sur le budget de l'État, puisque ce sont des dépenses permanentes du service public » (Trojette et Lombard, 2013).

auditionné pour le rapport. Pascal Berteaud, directeur général de l'IGN y rappelle que l'*open data* n'est pas un nouvel objectif pour l'IGN. En 2013, des jeux de données sont gratuitement disponibles à tous et les collectivités ont gratuitement accès à plusieurs référentiels. S'il rejoint la conclusion du rapport Trojette selon laquelle « ce serait mieux si toutes les données étaient ouvertes ! ». Il signale cependant que « la vraie question est quel sera notre modèle économique pour produire des données ? ». Il en appelle alors à une décision politique gouvernementale. Enfin, les jeux de données d'informations géographiques de l'IGN exigent une finesse de collecte et une actualisation constante. « Pour ce faire, aujourd'hui nous percevons une redevance qui couvre en partie ces dépenses ». L'absence de cette redevance signifie, dans l'immédiat une baisse de la qualité des données produites. Il envisage la possibilité de construire des services rémunérateurs pour l'organisme national mais les perspectives restent floues. C'est ce vers quoi Gaël Musquet, président d'OSM France semble encourager l'IGN. À sa suite, dans l'article de la Gazette des Communes, Gaël Musquet considère que « l'IGN doit revoir son mode de gouvernance » (Blanc et Fauvel, 2013). Il affirme être en accord avec la conclusion du rapport Trojette selon laquelle la prise en charge d'une part des coûts de production des données publiques par les ré-utilisateurs fait peser un risque important sur la pérennité du service public. Dans le cas de l'IGN il s'agit « d'accompagner » l'organisme vers un modèle ne nécessitant pas de redevance. Il considère que l'organisme ne dispose pas des moyens humains nécessaires pour garantir son service. Il juge que « l'agence vieillit » (Blanc et Fauvel, 2013) et ne s'est pas adaptée aux nouveaux usages. Face à ce constat, il propose à l'IGN de considérer la capacité d'édition des contributeurs à OpenStreetMap. Il termine en demandant à l'IGN « d'arrêter de [les] considérer comme des flibustiers, mais comme des partenaires ». Une réponse à ces remises en question est publiée le lendemain dans la Gazette des Communes. Pascal Berteaud écrit un papier d'opinion qu'il intitule *Pour en finir avec de fausses informations sur l'IGN !* Il accuse « certains opérateurs de la cartographie numérique » de chercher à dénigrer l'IGN dans le but de construire leur notoriété. Il souhaite corriger l'accusation d'un vieillissement de son organisme en soulignant que l'âge moyen, de 44 ans, est égal à celui de la fonction publique territoriale. Ensuite, considérant les reproches d'une incompatibilité de services entre l'IGN et les « nouveaux » besoins, il répond que l'IGN est constitué de laboratoires de recherche « internationalement reconnus » et actifs dans des domaines de nouvelles technologies (GPS). De plus, il explique que l'évaluation de la capacité de collecte est

bien supérieure à celle énoncée par Gaël Musquet et il ne répond pas directement à l'offre de partenariat du président d'OSM France mais sous-entend un manque de transparence dans la gestion de l'association, une gouvernance dictée par des « commanditaires anglo-saxons » (Berteaud, 2013) une licence « contaminante » (Berteaud, 2013). Enfin, Pascal Berteaud rappelle que le système actuel de financement de l'organisme s'explique par la demande de l'État que l'IGN assure 35% de ses dépenses.

Dans la construction de sa pensée et dans le cas particulier du champ scientifique, Pierre Bourdieu (1975) semble avoir exclu l'intérêt économique comme un des facteurs de la recherche de monopole scientifique. Dans le contexte actuel, les enjeux économiques sont pourtant constitutifs du rapport de forces entretenu entre OSM et certaines autorités scientifiques. En effet, le schéma traditionnel selon lequel les autorités scientifiques garantissent des procédés d'évaluation préalables à leurs publications, exige des fonds financiers importants. Ces fonds ne sont pas toujours garantis par les États. Dans le débat survenu autour de la publication du rapport Trojette, cet enjeu économique a bien été source de conflits mettant en question l'autorité des intervenants au débat[150]. La section 4.1 du rapport Trojette, intitulée *Les limites des modèles économiques retenus- La fin du monopole public dans l'élaboration de certaines données de référence*, présente le projet OSM comme concurrentiel à l'IGN. Il est précisé que si le monopole de l'IGN « s'érode progressivement, compte tenu de l'initiative de collectivités territoriales, des moyens de certains grands groupes mais aussi [...] des démarches ouvertes, participatives et collaboratives (« *Crowdsourcing* ») » (Trojette et Lombard, 2013, p.67) ce n'est pas le fait « d'un souhait de concurrencer les sources d'informations publiques, mais plutôt de la présence de barrières aux usages non professionnels, telles les redevances de réutilisation » (Trojette et Lombard, 2013, p.67). La pratique de collecte collaborative citée vise OpenStreetMap mais également Google. Il est remarqué que « le grand public, les collectivités territoriales, les entreprises mais aussi par plusieurs services de l'État, tous clients traditionnels de l'établissement public » (Trojette et Lombard, 2013, p.68), sont des utilisateurs de ces données. Ainsi, la concurrence est signalée comme réelle mais il est envisagé que ces

[150] On remarque d'ailleurs que l'encart de description du projet OpenStreetMap (p.70) du rapport Trojette cite le récit fondateur d'OpenStreetMap qui met en jeu Steve Coast et l'Ordnance Survey. On peut y lire : « Le constat de Steve Coast a été que l'agence cartographique publique de son pays, l'Ordnance Survey, conserve le droit de reproduction à son profit, alors qu'elle est financée par ses principaux utilisateurs, les contribuables britanniques. » (Trojette et Lombard, 2013).

sources de données puissent être complémentaires. On peut considérer, dans cette discussion, que la Cour des comptes accorde une légitimité au processus de création d'OSM. Enfin, rappelons qu'OSM peut être utilisé comme moyen de concurrence économique mais qu'il n'est pas lui-même un acteur économique. C'est l'usage d'OSM par un acteur tiers qui en fait un enjeu concurrentiel en matière d'intérêts économiques.

Si OSM n'est pas un acteur politique déclaré, sa présence forte dans l'espace d'Internet en fait un acteur politique effectif. Les conférences OSM et notamment l'intégration des présentations par des chercheurs, sont des indications de la mise en place progressive d'un message pédagogique et méthodologique plus clair. Mais le Libre implique inévitablement une forme marquée de vulnérabilité : seule la réduction régulière et manifeste de ses risques d'erreurs par la limitation de la vulnérabilité permettra à terme à OSM de devenir une autorité.

2.2.3. Vulnérabilité

Le mode de production des données OSM repose donc sur le processus de rectification. La méthode de rectification choisie est elle-même issue des concepts du mouvement du Libre selon lesquels la création itérative et ouverte mène au raffinement de l'objet produit. Le caractère itératif d'une production libre permet la rectification tandis que le caractère ouvert soutient une rectification par le plus grand nombre. Cependant, c'est ce caractère ouvert du projet qui lui confère une vulnérabilité liée au risque de dégradation de la base de données. En effet, la base de données est exposée au vandalisme, aux contributions erronées et aux manipulations involontaires. Ces trois types de dégradations sont à distinguer par leur ampleur et en fonction de l'intention du contributeur responsable de la nuisance. En effet, les erreurs de toponymies, de topologies et de catégorisations sont nombreuses. Cependant, considérant le mode de production elles sont identifiées par des procédés logiques mais également par le contrôle manuel, et sont corrigées progressivement. L'erreur et l'approximation sont donc admises comme normales, éthiquement acceptables de la part des contributeurs et solvables du fait de leur caractère ponctuel. Au contraire, le vandalisme en tant que dégradation volontaire et illégitime représente une menace pour le projet OSM en tant qu'acte volontairement malveillant.

La vulnérabilité est inhérente au système libre et contributif. Néanmoins, en balance, est opposé le postulat selon lequel le projet contributif peut réunir, par l'ouverture, un plus grand nombre de personnes soucieuses d'en améliorer la qualité que de personnes animées d'une volonté de nuire. Ce postulat, partagé par les contributeurs à OSM et par les contributeurs à Wikipédia, est analysé par Glassey (2006) qui prend l'exemple de Wikipédia. Il signale que cette pratique de la création repose sur deux paris qui nous semblent également concerner OSM. Le premier pari consiste à considérer « une contribution jugée positive [comme ayant] des chances de demeurer dans le corps de l'article alors qu'une contribution jugée « nuisible » sera détruite d'un seul clic par le contributeur suivant ». Ce premier pari repose donc sur le poids démographique constitué de la population des personnes bien intentionnées par rapport à la population qui ne l'est pas. Le second pari, identifié par Glassey (2006), est celui de l'efficacité de la remise en cause perpétuelle comme méthode d'amélioration substantielle du projet. Selon Glassey (2006), la pratique du consensus est conçue comme « un tamis servant à l'identification et à l'éviction des propositions trop déviantes ou extrémistes ». Dans le cadre de l'encyclopédie, il s'agit de rédiger un article qui, tout en s'appuyant sur la pluralité des points de vue, se rapproche au plus près de la vérité du fait relaté. Dans le cas d'OSM, c'est la finesse de description, la création d'une nouvelle clé ou la concordance de l'ajout d'un nouveau type d'objet avec les visées du projet qui est débattu jusqu'au consensus. Cependant, comme l'écrit Glassey (2006), « alors que paradoxalement des ressources considérables, sont dévolues à la préservation de la mémoire écrite des étapes antérieures, rien n'empêche, formellement, un nouveau venu de remettre en cause le produit d'un long processus de négociation ».

La vulnérabilité d'un projet libre et contributif est donc avérée. Cependant, si le projet OSM est exposé à la vulnérabilité vis-à-vis de dégradations, cela ne fait pas d'OSM une source instable. Un *commun,* tel qu'OSM, se développe et se maintient grâce à des mesures de protection qui sont appliquées en fonction de règles admises par le groupe. Ainsi, le risque existe si des mesures de protection ne sont pas mises en œuvre et dans le cas où ce rapport de force, entre contributions positives et contributions nuisibles, s'équilibrerait ou s'inverserait. Alors seulement, la qualité pourrait rapidement péricliter. Les mesures de protection doivent ainsi répondre à trois risques. Le premier risque est celui des erreurs de contributions ou de manipulations involontaires. Les deux autres risques sont des types de vandalisme : un vandalisme malveillant et un

vandalisme relevant du comportement individuel d'un contributeur qui ne se préoccupe pas des choix de modélisation établis par la communauté.

Les mesures de protection face aux erreurs et aux manipulations involontaires consistent en un ensemble d'outils de contrôle et de correction des données[151]. Le fonctionnement par la rectification est donc une mesure de protection *a posteriori* et repose principalement sur le poids démographique des contributeurs volontaires à l'amélioration des données. *A priori*, qu'elle soit volontaire ou non, une dégradation pourrait être traitée par les mêmes procédés de rectification. Les outils de contrôles sont les moyens les plus couramment mobilisés. En cas de dégradation conséquente, ce sera le *Data Working Group* (DWG) qui interviendra. Ce groupe de travail de la Fondation OSM est responsable des questions relatives aux données. Il s'attache à résoudre les conflits dans les cas de violation du droit d'auteur, de guerres d'édition, de vandalisme. Il dispose d'un pouvoir coercitif contre les contributeurs. Afin de s'assurer que l'application des règles soit fonction de l'ampleur de la dégradation, le DWG dispose des moyens suivants :

- Commenter publiquement les discussions relatives à un *changeset*;
- Adresser un message au contributeur, par la voie normale d'échange entre contributeurs ;
- Adresser un message au contributeur qui doit être lu par le contributeur avant que celui-ci puisse de nouveau éditer la base, ce blocage est visible publiquement ;
- Supprimer des notes offensantes ou des commentaires de *changeset ;*
- Bloquer un compte d'utilisateur jusqu'à 96 heures maximum ;
- Recommander à la Fondation OSM de bloquer définitivement un utilisateur.

Si ces règles visent une réponse à la mesure de la dégradation identifiée, il est précisé, à la page détaillant la politique de bannissement du site de la fondation[152], qu'il existe une exception à cette pratique graduée. On peut y lire que la parcimonie n'est pas réservée aux « comptes » qui contribuent manifestement dans le seul but de vandaliser OSM. Ces comptes peuvent être bloqués ou interdits sans délais. Cette règle ne s'applique pas non plus aux contributeurs qui ne répondent pas aux tentatives répétées de communication avec le DWG. Ils peuvent être bloqués jusqu'à ce qu'ils répondent. En outre, cela ne s'applique pas aux contributeurs qui violent le droit d'auteur, les droits

151 Exposés précédemment dans la section II.2.1.
152 Article *Ban Policy*, sur le site /wiki.osmfoundation.org, consulté en ligne en janvier 2019.

de base de données ou d'autres droits similaires. On comprend, dans cet exposé, les causes potentielles d'un bannissement, que la Fondation ressent, à propos du vandalisme, une inquiétude pour la base de données mais également pour le bon fonctionnement en groupe de contributeurs. Le non-respect des droits de licence, les disputes via le *wiki* ou les discussions autour des *changesets* ou des comportements de dérangement répétés relèvent du vandalisme.

En ce qui concerne la dégradation de la base de données, en 2012, Neis et *al.* considèrent le vandalisme comme un ensemble regroupant des actes intentionnels et non intentionnels. À la recherche d'un système de repérage d'actes de vandalisme qui soit rapide et utilisable par les contributeurs, il propose des règles caractéristiques du vandalisme dans OSM. Deux types de paramètres sont pris en compte, la réputation du contributeur et les caractéristiques de contribution. S'appuyant sur les causes, disponibles publiquement, qui ont menées le DWG à bloquer des contributeurs, ils proposent un ensemble de règles qui, si elles surviennent ensemble peuvent indiquer une dégradation. Ainsi, pour évaluer la rigueur de description attributaire, le *wiki* est utilisé comme moyen de vérification. Un *tag* ajouté à un objet qui ne serait pas documenté dans le *wiki* est un élément indicateur d'une possible dégradation. Également, un *changeset* essentiellement constitué de suppressions ou d'ajouts importants d'objets peut devenir suspect. En 2017, Neis reprend ces premières propositions et développe un outil en ligne, qu'il nomme *Find Suspicious OpenStreetMap Changesets*[153]. Simple d'utilisation l'outil signale des *changesets* dont les contributeurs sont peu expérimentés (nombre de jours d'activité), ont tendance à supprimer, modifier ou éditer des grandes quantités de données et ne commentent pas leurs modifications[154].

Cependant, une dégradation bien pensée peut ne pas présenter les caractères classiques du vandalisme. Son identification peut alors être impossible par procédé logique car la dégradation est d'ordre sémantique. Néanmoins, le vandalisme volontaire est une action « marginale » (Truong et *al.*, 2017) qui, pour être efficace, doit être visible de tous. Une action de dégradation des données doit donc viser des objets à forte symbolique. Le 10 août 2018 à 23h44, le contributeur MedwedianPresident décide de remplacer le nom de

153 Trouver des groupes de changement suspects dans la base OpenStreetMap.
154 Outil en ligne :
https://resultmaps.neis-one.org/osm-suspicious?country=-1&hours=12&mappingdays=-1&tsearch=review_requested%3Dyes&anyobj=t&comp=%3E&value=10&action=d&obj=n#2/13.9/9.3

la ville de New-York par un propos antisémite. Deux heures plus tard, le 11 août 2018 à 1h43 la correction est réalisée, le *tag* `name=New York` était de nouveau inscrit et le *changeset* précise : « *Reverted changeset 61555047 (https://www.osm.org/changeset/61555047) due to vandalism* ». Enfin, le 31 août 2018, le DWG bannit pour 19 ans le contributeur *Naoliv* pour cause de vandalisme et l'invite à contacter la Fondation[155]. Cet exemple est emblématique de la vulnérabilité du projet mais révèle également la capacité de réactivité des contributeurs et l'organisation d'un cadre réglementé. La force du poids démographique est efficace dans ces cas symboliques. Précisons que si l'action coercitive par le DWG participe d'une atmosphère organisée, il est toujours possible pour le contributeur de se créer un nouveau compte. C'est d'ailleurs ce que le contributeur en question a fait en ouvrant un compte le jour de son bannissement, sous le nom de *hahahahahhahahahahaha*. Par ce compte, il a vandalisé l'objet représentant la seconde avenue à New York en remplaçant son nom par une adresse URL d'une page de forum dans laquelle il explique ses motivations de « vandale ». Il a fallu quinze minutes pour qu'un membre du DWG corrige la donnée. Nous pouvons constater ainsi une réelle vivacité des contributeurs quant aux attaques qui concernent des objets à forte symbolique. De ce fait, en cas d'usage public et en ligne de la base de données, il semble plus cohérent, de s'assurer d'une mise à jour très régulière de la version de la base de données OSM utilisée plutôt que de mettre en place des traitements longs et qui resteront insuffisants pour l'identification de l'ensemble du vandalisme.

Alors que le vandalisme volontaire, symbolique et marginal, ne constitue pas un risque de dégradation important, un vandalisme plus insidieux, issu des comportements individualistes et délibérément indifférents au consensus, pose davantage problème. Deux situations répondent à cette description. La première s'explique par un malentendu entre services utilisant les données et services d'édition de la base de données. Lorsque la base de données est utilisée dans le cadre de services de cartographie ou de jeux mobilisant des données géographiques, il arrive que les moyens d'édition de la base OSM, en ligne ou par les applications, ne soient pas compris comme tel mais plutôt comme des moyens d'interactivité permettant de dessiner et qui serait conservé en local sur l'ordinateur ou le téléphone. Dans ces cas, le concept de la base

[155] Message adressé à Naoliv par le Data Working Group de la Fondation OSM en en août 2019 : https://www.openstreetmap.org/user_blocks/2141

de données libre n'est pas perçu par l'utilisateur qui devient vandale par méconnaissance. On peut lire aussi dans les échanges entre contributeurs avertis et contributeurs malgré eux que ces derniers pensaient qu'une modération empêcherait leurs éditions d'être enregistrées. C'est en partie ce qu'il s'est produit lorsque certains joueurs de Pokemon Go se sont aperçus que les données OSM étaient utilisées. En effet, il est présumé que, dans le jeu, certaines espèces de *pokemons* apparaissent à proximité de types d'objets géographiques spécifiques comme les immeubles, les lacs ou les voies piétonnes. Trouvant dans l'édition de la base de données une solution simple pour faire apparaitre des *pokemons* de types nouveaux dans leur entourage, des joueurs ont donc contribué des objets fictifs. Ces contributions relèvent du vandalisme au sens où l'action de contribution est volontaire est malhonnête. La pratique a cependant été mise en évidence par les méthodes classiques de détection d'incohérence. En effet, les joueurs bien que malhonnêtes n'étaient pas malveillants et leurs contributions n'étaient pas conçues pour être dissimulées. Des voies piétonnes non contiguës au réseau ou des surfaces intersectant d'autres polygones sont les principales pratiques relatives à ce jeu. Le second comportement individualiste menant au vandalisme caractérise certains contributeurs plutôt experts. On en trouve des exemples parmi les contributeurs bannis par le DWG. Ainsi, le contributeur *RB94*, d'abord bloqué pour des périodes temporaires de quelques jours, l'est définitivement après une dernière tentative de coercition temporaire : « tu continues à modifier la carte alors que les autres éditeurs de ta région ne sont pas d'accord avec ces modifications. Tu continues à changer les noms pour indiquer des détails (direction…) qui ne font pas partie des noms. Tu ne réponds toujours pas aux autres contributeurs qui te contactent. Nous avons maintenant bloqué ton compte ». Finalement il est bloqué de façon permanente par un membre du DWG. Dans un cas comme celui-ci, où le contributeur a pu à la fois créer de la donnée utile et compatible avec les pratiques communes et à la fois les détériorer, c'est aux contributeurs d'identifier les dégradations. Les contributeurs concernés, c'est-à-dire, essentiellement situés en Ile-de-France et contributeurs de données de transport ont donc mis en place un challenge, via l'outil *MapRoulette,* dans cet optique (fig. 23).

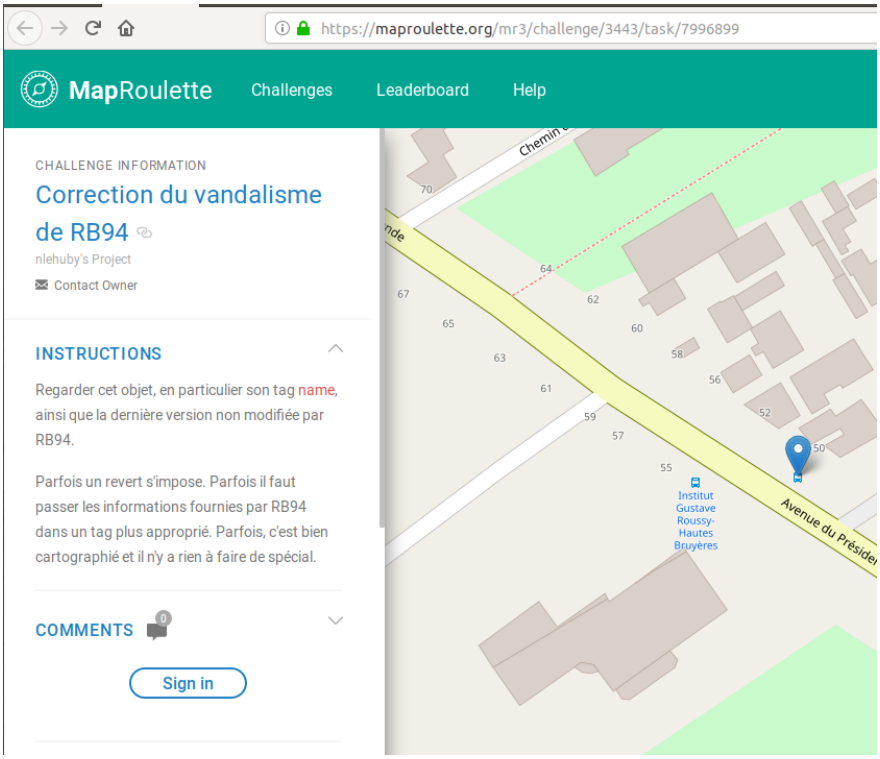

Figure 23 Challenge MapRoulette répertoriant les objets dont RB94 est le dernier modificateur, capture d'écran sur le site MapRoulette

Le vandalisme prend des formes diverses, très complexes à identifier tant les motifs sont nombreux. Neis et *al.* (2012) considèrent les contributeurs débutants comme plus probablement vandales mais nous avons vu que le vandalisme peut également être pratiqué par un contributeur expérimenté qui s'isole. Les outils logiques servent à l'identification de toutes les formes de dégradations. Cependant, afin de se protéger du vandalisme symbolique mais également de comportements individualistes, il importe que les contributeurs développent de nouveaux moyens de communication. Dans les cas où des débutants ont pu ne pas comprendre l'immédiateté de leur modification, il s'agirait d'identifier par quel outil l'édition a été réalisée et si les phases explicatives ne pourraient pas être plus explicites. En ce qui concerne le mépris du consensus par des contributeurs expérimentés, c'est le poids démographique qui représente la solution. Car si le nombre permet d'identifier rapidement une dégradation, il s'agit de proposer des solutions simples (comme MapRoulette) pour participer à leur correction.

2.3. OSM, l'incompris

À la question de la capacité pour OSM de devenir une autorité reconnue par les praticiens du domaine de l'information géographique, nous avons répondu doublement. D'une part, OSM n'est pas une institution clairement identifiée et cette réunion de contributeurs n'a pas pour visée de répondre aux demandes spécifiques d'un pouvoir public, au service des citoyens. D'autre part, OSM est une réelle alternative dans les contextes où les instituts nationaux sont inexistants ou ne disposent que de faibles moyens. Enfin, en contribuant sur des thématiques qui motivent le contributeur, même si parfois à faible portée en termes d'audience, les contributeurs créent une base de données riche en informations alternatives. Questionner le statut d'autorité d'OSM dans la sphère de l'information géographique nous permet d'étendre la question de la fiabilité à l'ensemble du projet OSM plutôt qu'uniquement à la base de données. De ce fait, on pose la question de la fiabilité sur un temps plus long que celui de l'état actuel de la base de données. Nous nous interrogeons alors sur les perspectives possibles pour le mode de production de la base de données contributive. Lorsque l'on pose la question de la fiabilité de l'ensemble du projet OSM, c'est-à-dire sur le processus de création de la base de données OSM et le résultat de ce processus, il nous semble important de reconnaître le fort potentiel d'évolution du projet (3.2. Ce que le projet peut devenir).

Cependant, il apparaît que les évaluations de qualité attribuent à OSM un rôle que le projet ne prétend pas porter. En effet, une des erreurs qui mène souvent à considérer que la qualité d'une source est mauvaise consiste à attribuer à celle-ci un rôle ou un usage qu'elle n'a pas. Dans le cas d'OSM, cela s'explique en partie par le fait que le projet lui-même ne précise pas ses limites. Si OSM ne se défini ni un rôle ni un usage, il convient de rappeler ceux auxquels le projet ne peut prétendre.

2.3.1. Ce que le projet n'est pas

OpenStreetMap ne peut pas être considéré comme une institution. Le premier élément explicatif est qu'OpenStreetMap suit une organisation largement horizontale. En effet, la Fondation OSM qui peut être l'interlocuteur considéré comme représentant d'OpenStreetMap porte seulement deux responsabilités : la publication des données et leur protection juridique. La Fondation ne garantit pas de processus de contrôle et ne

prétend pas être apte à contrecarrer le vandalisme. On ne peut donc pas considérer la Fondation comme la représentante d'OpenStreetMap en tant qu'assurant un système de production fiable. Il pourrait être envisagé de considérer les associations locales comme des représentations d'OpenStreetMap à l'échelle des territoires couverts par leurs membres. C'est en partie ce que l'on a pu observer à travers le rôle pris par OpenStreetMap France dans le projet pour la BAN, ou son usage par Etalab, ou encore dans l'entente avec l'IGN pour la signature d'un contrat d'exclusivité de l'usage de certaines données institutionnelles comme le cadastre et la BD Ortho. Cependant, l'usage de la base de données OpenStreetMap concerne souvent son intégralité, lorsqu'elle est mobilisée pour de la cartographie en ligne. Tous les territoires couverts ne disposent pas d'une association locale apte légalement à signer des contrats ou à participer à une activité d'intérêt public. Le second élément explicatif est que cette organisation horizontale est volontaire tandis qu'aucun groupe, fondation ou association d'OSM ne prétend porter l'autorité sur les autres groupes. De ce fait, puisqu'aucune prétention à être reconnue comme une autorité ne se dégage et qu'aucune organisation de groupe ne vise à atteindre un service d'intérêt public, OpenStreetMap ne peut être considéré comme une autorité du domaine de l'information géographique au sens traditionnel.

Ce contexte dépourvu de garanties explique le principal défaut de qualité de la base de données OpenStreetMap, c'est-à-dire son hétérogénéité spatiale et attributaire. Les premiers travaux scientifiques relatifs à la qualité des données produites par des contributeurs ont expliqué que le caractère indicatif et non prescriptif des spécifications du projet OSM, écrites sous la forme d'un *wiki* en ligne, est un facteur d'une qualité hétérogène des attributs (Girres, Touya, 2010). Les *tags* les plus fréquents sont votés tandis que d'autres sont en usage sans avoir été votés. On note une caractéristique supplémentaire dans la définition de ces spécifications : celles-ci sont souvent rédigées après qu'un certain nombre d'objets a été créé selon le même modèle. C'est ce que met en évidence le graphique ci-contre (fig. 24), conçu par Martin Raifer et Franz-Benjamin Mocnik en 2016[156].

156 Outil en ligne : http://projects.mocnik-science.net/osm-vis/osm-tags-wiki-vs-osmdata/, consulté en mai 2019.

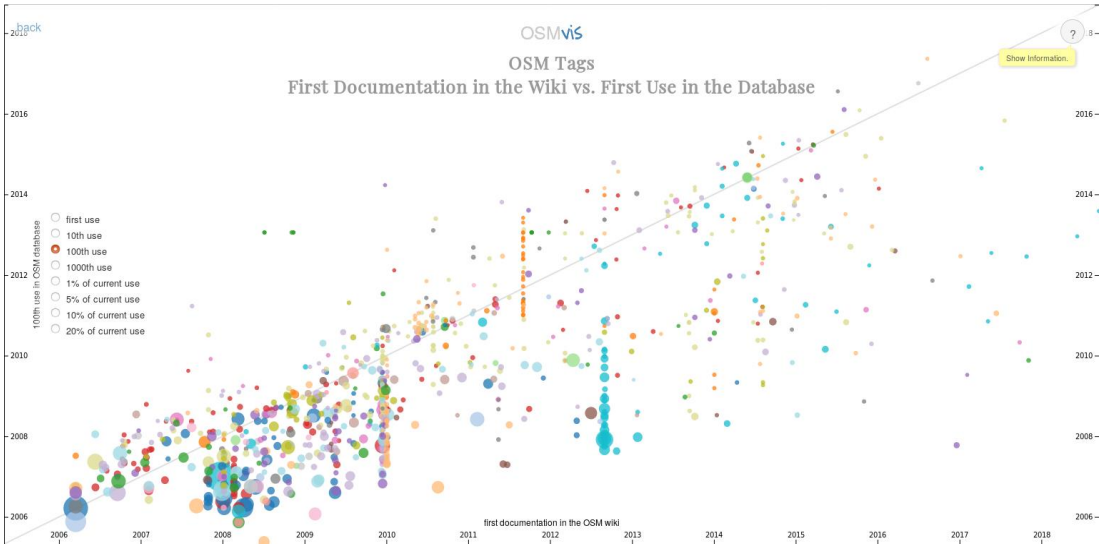

Figure 24 En abscisse : la première documentation du tag ; en ordonnées : la centième utilisation dans la base de données OSM du tag

Le constat de l'hétérogénéité de la base de données est immédiat. Certains travaux de recherche ont donc posé la question de la forme de répartition que prend cette hétérogénéité. D'après Hacklay (2010) et Zielstra et Zipf (2010), la couverture relativement homogène du territoire dans OSM diminue depuis les centres villes vers les zones rurales. En 2010, Hacklay explique que le processus de contribution volontaire n'est pas construit sur le même principe que celui qui guide les institutions responsables de la création d'informations géographiques. Celles-ci visent à soutenir une justice sociale par une couverture géographique égale sur l'ensemble du territoire dont l'institution est responsable. Hacklay (2010) mobilise ses mesures de l'exhaustivité des jeux de données centrés sur l'Angleterre en 2008 comme un indicateur de justice sociale. Il rappelle l'engagement des instituts nationaux de cartographie et des services postaux nationaux à fournir une couverture complète du pays, indépendamment de l'isolement d'un lieu ou de la situation socio-économique de ses habitants. La forte hétérogénéité spatiale qu'il identifiait en 2010 l'a mené à conclure que le projet OpenStreetMap n'était pas inclusif. Il montre que les contributeurs évitent les lieux socialement marginaux (identifiés par les relevés statistiques de 2007 sur les *Super Output Areas*). En 2012, Goodchild et Li rejoignent l'inquiétude exprimée par Haklay relative au risque d'injustice sociale suite à une contribution motivée par le seul intérêt personnel des contributeurs. Dans leur article, Goodchild et Li ne considèrent pas que la couverture régulière d'un territoire puisse être réalisée par des contributeurs libres de créer de l'information sur ce qui leur plaît.

Intégrant l'information géographique volontaire dans un courant plus vaste de publications d'informations sur Internet, les auteurs considèrent que le caractère géographique de cette information ne permet pas aux contributeurs de proposer des données de qualité. Goodchild et Li présentent une réalité qui distingue les projets contributifs libres des instituts nationaux responsable d'une égalité entre les individus du territoire qu'ils observent. Ils qualifient certains lieux comme « obscurs » aux yeux des contributeurs. Ces lieux « obscurs » sont peu peuplés, ne sont pas connus dans l'histoire et ne présentent pas d'objets géographiques qui seraient attrayant à renseigner dans OSM. De ce fait et, selon eux, ces lieux de l'ombre échappent au cadre d'application de la loi Linus[157]. Ainsi, Goodchild et Li n'acceptent-ils pas l'hypothèse qui consiste à proposer la *loi Linus* comme une mesure d'assurance de la qualité des données géographiques volontaires. En outre, ceux-ci considèrent que le nombre d'éditions ne constitue pas un paramètre prouvant la qualité de l'information. Ils s'appuient pour cela sur l'existence des guerres d'édition. En 2012, Mooney et Corcoran signalent cette réalité de la contribution dans OpenStreetMap que sont les guerres d'édition. Du fait de l'existence des guerres d'édition, un nombre important d'éditions d'un objet ne peut être considéré par Goodchild et Li comme un paramètre suffisant pour juger de la qualité d'un objet. Cette affirmation est un peu rapide. Depuis 2012, plusieurs chercheurs, s'inspirant de la Loi Linus ont proposé des procédures qui permettent de prendre en compte cette réalité. Ainsi, selon le type de contribution, qui relève soit de l'enrichissement itératif soit d'un aller-retour entre des descriptions semblables, les indicateurs de qualité jugeront ce paramètre comme un indice négatif et parfois comme un indice positif de la qualité de la donnée (Keßler et De Groot, 2013).

L'aspect hétérogène de la base de données nous mène aujourd'hui à considérer OSM comme partiellement moteur d'originalité en faveur de thématiques locales mais

157 Pour que la Loi Linus s'applique, il est d'abord de rendre accessible le projet OSM au plus grand nombre. Progressivement, le Géoweb rend davantage accessible des techniques qui nécessitaient auparavant une expertise. Comme le l'observent Vienne et *al.* (2017), « L'analyse des géoréférencements numériques dans les territoires de densités intermédiaires sur Facebook révèle une dynamique d'appropriation des ressources territoriales et de reconnaissance d'une localité numérique vécue. L'intégration des outils numériques dans le quotidien d'une partie des habitants des territoires dits périurbains accompagne une transformation de la représentation du territoire habité. Le changement paradigmatique du numérique affecte de manière sensible ces territoires des périphéries métropolitaines par une réappropriation progressive de son environnement proche ». Partant de ces résultats, relatifs à la dimension territoriale des pratiques de sociabilité numérique, issus de l'observation des occurrences relatives à 1935 lieux dans une zone de 250 communes françaises, nous posons l'hypothèse que la participation au projet OSM peut mener le contributeur vers un sentiment d'appropriation de son territoire. Ainsi, si la pratique de contribution à OSM est motivée par des facteurs divers, peut-on supposer qu'aucun lieu ne sera obscur à cette dynamique qui rejoint les espaces de sociabilité, virtuels et physiques.

également comme ne supportant pas systématiquement une justice sociale et spatiale. Pourtant, le principe d'un support à une justice sociale par l'homogénéité de couverture en informations géographiques actualisées est fondateur des instituts nationaux de cartographie.

Compte tenu de ce constat d'hétérogénéité qui repose sur le modèle même de construction de la base de données OpenStreetMap, il s'agit d'évaluer la qualité des données en fonction de leurs usages et des alternatives à son usage. Les instituts nationaux de cartographie ont la responsabilité de maintenir un nombre restreint d'informations nécessaire à la conduite de politiques de gestions des pays. Ceci n'est pas le rôle que s'est donné OpenStreetMap. **La construction d'OpenStreetMap est issue d'un désaccord avec le principe de non gratuité des données géographiques.** Cependant, ce que le projet est devenu dépasse largement ce conflit. **OSM est un espace de partage d'informations et de conception de ces informations libres et donc propice à l'originalité.**

2.3.2. Ce que le projet peut devenir

Nous avons décrit les limites qu'il s'agit de définir lorsqu'on pose la question des possibilités offertes par le projet OSM. Ces limites sont importantes et l'hétérogénéité comme la vulnérabilité des données sont des limites immuables car inhérentes au projet. Cependant, le projet est dynamique et des perspectives s'ouvrent avec son installation dans la durée. Nous dégageons trois perspectives principales relatives à la fiabilisation de la production par les contributeurs à OSM.

La première perspective est celle d'une **poursuite de la fiabilisation de la base de données et du cadre de production** par l'application du procédé de rectification continue. La tendance à l'amélioration par les contributeurs, vers davantage d'exhaustivité, s'est d'abord exprimée à travers des thématiques peu développées par les producteurs traditionnels. Ainsi, par exemple, en 2011, l'écart kilométrique entre le réseau routier d'OSM et celui de *MultiNet* de TeleAtlas, était de 9 % en faveur de la base *Multinet* pour les routes catégorisées comme praticables par les véhicules motorisés. Cependant, le réseau allemand OSM présentait 27 % de kilomètres supplémentaires en comparaison à la même base (Neis et *al.*, 2012). Cela s'explique

par des thématiques différentes entre ces deux producteurs. TeleAtlas était un producteur spécialisé dans la production de données pour la mobilité routière (comme TomTom qui en est propriétaire depuis 2008) et se souciait davantage de la qualité topologique de son réseau (un exemple de la cohérence logique très hétérogène dans OSM).

Au contraire, les contributeurs au projet OSM se sont rapidement concentrés sur le réseau piéton en ville, puis sur les chemins pédestres et le réseau de voies cyclables (Zielstra et Zipf, 2010, Neis et *al.*, 2012). On note ce fort intérêt pour les voies piétonnes dans l'ordre d'apparition des valeurs de la clé `highway`. Le *tag* `highway=living_street` est le premier *tag* précisant la catégorie de voie dessiné, à être documenté dans le *wiki* (juillet 2007), il est utilisé pour qualifier un objet de la base de données en août 2007. Ensuite, en novembre 2007, les voies piétonnes (`highway=pedestrian`), les voies cyclables en site propre (`highway=cycleway`) et les voies autoroutières sont contribuées. Les *tags* `highway=steps`, `highway=footway`, `higway=service` et `higway=bridleway` sont utilisées pour les premières fois au début de l'année 2008. Les *tags* décrivant des catégories intermédiaires de la hiérarchie du réseau comme les voies primaires et secondaires n'apparaissent que plusieurs mois plus tard[158]. En outre, bien qu'il ait été constaté que le réseau OSM dépassait, dès 2010, en quantité le réseau publié par TomTom, leurs usages n'étaient cependant pas comparables. D'abord, le réseau routier présent dans OSM restait inférieur de 9 % à celui de TomTom. Il le dépassait par son réseau pédestre, supérieur de 31 % à celui de TomTom. On peut considérer, selon l'usage, un niveau d'exhaustivité plus ou moins utile. En ce qui concerne l'exhaustivité attributaire, 16 % des voies routières ne présentaient ni de nom ni de numéro de référence. Cela a un fort impact sur l'usage pour de la navigation routière. Neis et *al.*, (2012) concluent, après expérimentation, que le réseau routier d'OSM est inutilisable directement pour le calcul d'itinéraire sans traitement de la donnée. Bien que la pratique exige de contribuer dans le respect de la topologie des objets les uns par rapport aux autres, les auteurs signalent des erreurs comme des tronçons non joints, des objets dupliqués ou des croisements non marqués par un nœud. A ce défaut de cohérence logique, ils associent un défaut d'exhaustivité attributaire. En

158 Outil en ligne : http://projects.mocnik-science.net/osm-vis/osm-tags-wiki-vs-osmdata/, consulté en mai 2019.

effet, l'attribut indiquant les interdictions de tourner était cinq fois plus présent dans la base de données de TomTom en comparaison à la base OSM. Un exemple notable de contribution est la participation de l'entreprise Mapbox au développement d'outils d'édition. L'entreprise MapBox, largement tournée vers des services de cartographie embarquée, a ainsi développé un certains nombres d'outils de correction topologique du réseau routier. *To-fix* est un outil de MapBox, installable en *plugin* dans JOSM qui signale une liste d'intersections entre deux tronçons possibles et pour lesquelles la donnée peut être à corriger. Comme nous l'avons vu, la pratique de la rectification s'applique aussi bien aux données qu'à la folksonomie et à sa documentation. En 2010, Girres et Touya analysent le cas de la qualité insuffisante des tracés de côte dans OSM, par rapport à la base de données TOPO de l'IGN. Ils suggéraient à ce moment-là le besoin pour OSM de définir des spécifications strictes, accompagnées de contraintes dans la création de données.

Le tracé de côte est un cas exemplaire des évolutions dans la réflexion de la modélisation par la communauté OSM. Cette réflexion s'accompagne d'une documentation écrite rendue nécessaire par l'extension de la communauté. En septembre 2009, date à laquelle la page du *wiki* relative au *tag* `natural=coastline` était créée, on considère que la discussion était en cours sur la façon dont il convenait de tracer le littoral lorsque la mer et la rivière se confondent dans un estuaire[159]. Cette discussion avait commencé deux ans auparavant sur la page de proposition de création du type d'objet : « Tidal Rivers ». On trouvait les interrogations suivantes : Si nous ajoutons une clé pour la marée ? Où s'arrête la marée ? Où s'arrête la rivière ? [160]. C'est finalement en mars 2014 que la définition de la clé tidal (marée) commence à être précisée dans le *wiki*. En 2016, l'usage de cette clé est clairement défini en déclarant une amplitude : « La zone dans l'amplitude des marées se situe entre la hauteur moyenne de pleine mer et la hauteur moyenne de basse mer. La zone est signalée par le *tag* `tidal=yes` »[161]. La hauteur moyenne de pleine mer est également employée pour dessiner le trait de côte. La page relative au *tag* `natural=coastline` est largement détaillée dans ses recommandations de contribution, appuyées (*Tag:natural=coastline,*

159 « *There is an ongoing discussion about how we should tag the coastline as it goes into an estuary and becomes a river* ». (Tag:natural=coastline, 2009).
160 « *If we add a tidal key, where does a river stop being tidal then?* » Proposed features/Tidal Rivers. (2018). Commentaire du contributeur *Emai* en janvier 2008.
161 « *Area in the tidal range between mean high water spring (natural=coastline) and mean low water spring is tagged with tidal=yes* » (Key:tidal, 2019).

2018). On voit donc que les contributeurs au *wiki* ont été conscients des indécisions dans la production, à l'origine d'erreurs de dessin du tracé des côtes, lorsque le contributeur n'a pas lui-même connaissance de ces indécisions. La thématique de la côte est un exemple qui permet également d'aborder le rapport à la qualité des données du point de vue des contributeurs expérimentés et impliqués dans la définition du modèle par leur contribution au *wiki*. En septembre 2009, le contributeur Markus B écrit que « d'un point de vue réaliste, la grande majorité des chemins portant le *tag* `natural=coastline` seront basés sur des données Prototype Global Shoreline (PGS) pendant une longue période encore et certaines seront corrigées à partir des images aériennes Yahoo ! (niveau de marée variable), mais si vous avez l'occasion de corriger les données pendant une sortie de terrain, c'est la limite de la marée haute qui est à relever »[162]. Sept ans plus tard, en janvier 2016, le contributeur Sanderd17 notait que « le tracé de côte a été, pour l'essentiel du tracé mondial, importé à partir de données générées à très petites échelles, les Prototype_Global_Shoreline (PGS). Les tracés inexacts peuvent être identifiés principalement par leur forme en dents de scie, alors qu'en réalité les côtes présentent la plupart du temps des courbes douces. Dans de nombreuses parties du monde, le tracé de côte a été amélioré, soit grâce au dessin à partir d'images aériennes, soit remplacé par de meilleures données vectorielles »[163]. Le niveau d'exigence attendu des contributeurs et par les contributeurs eux-mêmes est limité par la disponibilité en sources d'informations et par les moyens techniques de collecte. Ceux-ci devenant de plus en plus accessibles aux contributeurs, la précision des critères a été améliorée. Désormais, les contributeurs doivent[164] s'assurer que chaque chemin, constitutif du tracé de côte mondial, soit joint à chacun de ses voisins. En 2018, cet objectif n'est pas encore atteint. C'est pourquoi deux contributeurs Jochen Topf et Christoph Hormann proposent un service de téléchargement[165], externe à la

162 « *Realistically speaking the vast majority of our natural=coastline ways will be based on Prototype Global Shoreline (PGS) data for a long time to come, and some will be corrected based on Yahoo! Aerial Imagery (variable tide level), but if you get the opportunity to fix up the data while out surveying, then the high tide position is the one to go for*». (Tag:natural=coastline, 2009).

163 « *The shoreline data for most parts of the world was originally generated from a low-resolution data using the [[Prototype_Global_Shoreline | PGS]] import and other imports. Inaccurate shorelines can be seen mostly in a sawtooth shape, whereas in reality the coastlines are mostly gentle curves. In many parts of the world the coastline ways have subsequently been improved either with tracing from aerial imagery, or replaced by better vector data*. » (Tag:natural=coastline, 2016)

164 Pour chaque traduction des recommandations, on remarque des préconisations fortes (« *you must* ») qui encouragent le lecteur-contributeur à suivre ces recommandations, dans la mesure du possible le plus souvent (« *you should* »). On remarque également l'utilisation inhabituelle d'une forte graisse pour signaler certaines recommandations essentielles.

165 http://openstreetmapdata.com/data/coast

base de données OpenStreetMap, des tracés de côtes. Ces tracés sont traités par les auteurs, assurant des objets géographiques utilisables dans des SIG sans erreur topologique.

En 2017, Franz-Benjamin Mocnik, Alexander Zipf et Martin Raifer proposent des perspectives quant à l'évolution possible de la folksonomie d'OpenStreetMap. Nous sommes appuyés sur leur analyse de l'évolution de la folksonomie, de 2004 à 2017, pour avancer que les contributeurs suivent un processus de rectification indéfinie. De 2004 à 2017, les chercheurs ont distingué trois phases :
- entre 2004 et 2007, le projet est en construction et les premières clés de la folksonomie comme `highway` et `amenity` se développent sans profonde documentation,
- entre 2008 et 2009, la documentation est améliorée pour les clés principales,
- entre 2010 et 2017, la documentation tend à être entièrement complétée, le nombre de clés et de valeur augmente ; les chercheurs estiment que la folksonomie se raffine.
Deux phases prédictives sont proposées par les trois chercheurs :
- entre 2018 et 2031 la folksonomie se développera toujours vers une granularité plus fine mais atteindra une forme de stabilité avec la création de nouvelles valeurs et non de nouvelles clés ;
- à partir de 2032, le renouvellement de la folksonomie perdurera. Cependant, les nouveaux *tags* viendront remplacer d'anciens qui se présenteraient comme désormais inadaptées.

L'extrapolation présentée ci-dessus, à partir des tendances de contribution à la base de données et à la documentation, prévoit donc une poursuite de la rectification de la folksonomie par les contributeurs.

Un dernier élément constitutif de cette amélioration continue est la prise en compte des facteurs de motivation à la contribution. La motivation constante et continue des contributeurs chevronnés est un facteur de fiabilisation. Également, dans le contexte d'une méthode de création qui repose sur le débat de points de vue différents, la participation de nouveaux contributeurs, sur des territoires peu cartographiés ou d'appartenance sociale et professionnelle différentes des contributeurs actuels pourrait être un facteur de fiabilisation. Il est intéressant de noter que des moyens d'entretien de cette motivation se développent. Dans une perspective de spécialisation des contributeurs sur des thématiques de plus en plus finement décrites, des outils répondent

aux nouveaux besoins. Ainsi, les applications mobiles, visant à une contribution toujours plus simplifiée et en direct durant la promenade comme *OSMContributor*, *StreetComplete*, *OSM Go !* Ou encore *Keypad - Mapper* en sont des exemples récents et utilisés. De même, de nouveaux moyens de contribution visent à s'adresser à de potentiels nouveaux contributeurs. Afin de susciter un intérêt et un engagement pour la contribution, des jeux en ligne ont été conçus. Ces jeux visent à mener un novice ou un contributeur se désengageant, à réaliser une tâche par la ludification de cette tâche. En 2012 déjà, l'outil de correction d'erreurs en ligne MapRoulette était publié. Le site Web propose des challenges sur l'ensemble des données et le contributeur acquiert des points et se trouve classé. Déclarée comme un jeu, l'application Kort, initialement publiée en 2012, puis republiée en 2017 est de nouveau non maintenue. Comme MapRoulette, des tâches sont proposées aux contributeurs qui gagnent des points à leur réalisation. La présentation du jeu dans le *wiki* décrit Kort comme la mise en œuvre expérimentale du concept de *ludification* au service de l'accroissement de la motivation des contributeurs. En effet, cette application est un résultat issu du travail de thèse en informatique de Stefan Oberbolz à L'École technique de Rapperswil (Hunziker et Oderbolz, 2012). La diversité des applications ludiques ou simples, permettant la contribution de quelques *tags*, est représentative à la fois d'une spécialisation des contributeurs (transport, cyclisme, équipement, graphe routier) et des tentatives d'ouverture aux contributeurs débutants ou peu impliqués.

La deuxième perspective proposée ici, consiste à voir le projet OSM comme une **source de renouvellements méthodologiques**. Ces renouvellements concernent différents domaines relatifs à l'information géographique. D'abord, OSM présente un contexte de production original. Ce mode de production est propice à la créativité grâce aux interactions sociales et à la liberté d'expression. Cette liberté peut concerner un intérêt thématique particulier et une expressivité singulière en termes de modélisation de la réalité. La méthode du Libre offre des perspectives de production en groupe, via les outils numériques pour une expressivité aux contraintes limitées. Le cadre de production libre offre la possibilité de produire de l'information adaptée à la situation locale. Il n'y aura alors pas de raison de contribuer de façon identique dans d'autres zones. Ces thématiques sont très rares dans une base de données de référence, à l'exception de spécificités partagées par un pays entier comme certaines activités commerciales ou touristiques. Ensuite, nous avons pu constater que la communauté est

constituée de contributeurs à la base de données mais également de contributeurs à l'infrastructure technique. On peut citer les outils de travail collaboratif comme le *Task Manager* qui consiste à couvrir un territoire d'un ensemble de tuiles afin que chaque utilisateur se charge d'une parcelle. Dans la même perspective, d'une aide à la création contributive, l'Api OSM (qui a connu de nombreuses améliorations depuis 2006) est la démonstration de la possibilité technique de l'enrichissement d'une base de données en direct par plusieurs opérateurs ne communiquant pas nécessairement. Par ailleurs, l'existence de cette base de données a inspiré des innovations techniques dans l'optique de l'utiliser. Citons le développement de l'API *Mapbox GL*, largement utilisée pour la cartographie vectorielle ou la création d'autres projets ouverts nécessitant de l'information géographique comme Mapillary. Signalons également l'intérêt que représente le caractère ouvert de la base de données dans une pratique scientifique qui vise la reproductibilité des résultats mesurés à partir de données géographiques. Enfin, en parallèle, l'existence de la base de données OSM a été une opportunité de mettre en place des projets innovants. En cartographie, la liberté d'utilisation de la base de données a encouragé une plus grande diversité de représentations cartographiques centrées sur des thématiques (OpenSnowMap[166]) ou artistiques (Stamen[167], Makina Corpus[168]). Les circonstances de ces innovations reposent sur le statut libre du projet OSM. Ainsi, la méthode de production OSM est-elle une perspective de renouvellement méthodologique en tant que moteur pour l'innovation.

La troisième perspective, est fortement associée à ces nouveautés méthodologiques et consiste à considérer **OSM comme une source unificatrice** d'un point de vue technique et thématique. En effet, OpenStreetMap tend à s'intégrer dans le système de *linked data*. La perspective de l'interopérabilité des données OSM avec Wikipédia et Wikidata est en cours de réalisation. Une perspective plus large consiste à envisager l'interopérabilité de la BD OSM avec des bases produites par des institutions géographiques et statistiques. Comme nous l'avons vu, la vulnérabilité et l'hétérogénéité peuvent être acceptables selon les usages. Cependant ces deux aspects ne permettent pas de considérer OSM comme une source unique et institutionnelle. Nous proposons donc de considérer OSM comme une base pouvant accueillir

[166] http://www.opensnowmap.org/, consulté en mai 2019
[167] Carte *Watercolor* par Stamen :http://maps.stamen.com/watercolor/#12/37.7706/-122.3782, consultée en mai 2019
[168] Carte imitant l'esthétique du dessin au crayon à papier par Makina Corpus : http://dessine-moi-une-ville.makina-corpus.net/#18/43.60070/1.44236, consultée en mai 2019

largement des données institutionnelles sans que les garanties assurées par ces instituts ne soient dépréciées. En 2011, Dennis Zielstra et Hartwig Hochmair, mettent en évidence que, en ce qui concerne les données relatives à l'État de Floride, le réseau routier, les zones rurales, est plus important en comparaison à celui des deux grands producteurs de données que sont NAVTEQ et TeleAtlas. Les chercheurs expliquent cette spécificité dans la couverture du réseau par l'import massif du réseau de la base TIGER. En effet, ceux-ci soulignent l'usage dans le domaine agricole du réseau TIGER. La prise en compte des chemins d'exploitation dans la comparaison est la principale raison de cet écart. Ce résultat est inattendu car il contredit le schéma de contribution identifié jusqu'en 2011, via les études sur les espaces européens. En outre, le contraste avec la contribution européenne était accentué par une plus faible densité des données OSM dans les zones urbaines en comparaison aux producteurs commerciaux. Ainsi, Zielstra et Hochmair (2011) mettaient-ils en évidence une hétérogénéité contradictoire aux schémas de contributions habituellement identifiés.

Enfin, en lien avec le potentiel d'innovation, la base de données OSM est une source de données pour la recherche. On peut citer les études suivantes: «*OpenStreetMap data for alcohol research: Reliability assessment and quality indicators* » (Bright et *al.*, 2018), *Calculate travel time and distance with OpenStreetMap data using the Open Source Routing Machine (OSRM)* (Huber et Rust, 2016) ou encore « *Analyzing the effect of earthquakes on OpenStreetMap contribution patterns and tweeting activities* » (Ahmouda et *al.*, 2018). L'enrichissement de la base de données OSM par la mise à disposition des données d'autorité, garantissant par ailleurs la poursuite de leurs services traditionnels, offre une perspective pour le perfectionnement de l'interopérabilité entre les bases de données libres et pour la recherche plus largement.

Conclusion du chapitre 2

La première section de ce deuxième chapitre a été l'occasion d'exposer l'évolution des travaux scientifiques relatifs à l'évaluation de la qualité des données. La question posée est essentielle et permet de comprendre dans quelle mesure les données de la base OSM sont utilisables pour des usages cartographiques ou pour le calcul d'itinéraire. Comme cela a été noté, les six paramètres classiques d'évaluation de la qualité de données géographiques présentent des limites pour évaluer la qualité des données géographiques lorsqu'elles sont volontaires et libres. Les méthodes classiques d'évaluation de l'exactitude de position, l'exactitude attributaire, la cohérence logique et l'exhaustivité, sont des paramètres construits selon le principe qu'un jeu de données est homogène dans sa modélisation. En appliquant ces méthodes, on considère que les spécifications définies à la création de la base de données seront respectées. Par ailleurs, les évaluations mettent parfois en évidence que les spécifications sont incomplètes, imprécises ou inadaptées, laissant penser à un défaut de qualité général. Cette logique ne peut s'appliquer directement à OSM, sauf à considérer la base de données OSM comme étant structurellement de mauvaise qualité. Dans le cas où on souhaiterait dépasser cette limite et percevoir l'absence de contraintes logiques comme un nouveau mode de production, il est nécessaire de compléter les méthodes d'évaluation des données par des évaluations spécifiques aux besoins.

L'évaluation de la qualité des données repose également sur la compatibilité d'une source de données diffusée avec le besoin, les connaissances et les moyens techniques de l'utilisateur. En effet, en se démocratisant, les systèmes d'informations géographiques ont cessé d'être utilisés uniquement par leurs concepteurs. De nouveaux producteurs d'informations géographiques et de nouveaux outils de manipulations d'informations géographiques accessibles (en coût et en connaissances informatiques) ont permis un usage plus large des informations diffusées. Suite à ce constat, les chercheurs ont souligné l'importance du choix d'une source de données adaptée à l'usage qui en sera fait : « En s'éloignant de leur source [les données] risquent d'être utilisées pour des usages imprévus, pour lesquels elles seront peut-être inappropriées. Les utilisateurs leurs trouveront donc des défauts pour ce nouvel usage » (Rouet, 1991, p.133). L'exigence de l'évaluation de la qualité de l'information s'est accentuée avec l'apparition de nouveaux modes de production d'informations menant à la création

d'informations libres. Cette qualité est elle-même relative au besoin du cartographe. Ainsi, selon le besoin du cartographe, une source ou un ensemble de sources seront préférés pour leur système de référence, pour les garanties commerciales qui accompagnent leurs usages ou parfois pour l'accessibilité à l'information.

Les premières publications scientifiques, à partir de 2009, visant à évaluer la qualité des données d'OSM ont rendu compte des premiers succès de ce nouveau mode de production. Les données produites par les communautés de France, d'Allemagne et du Royaume-Uni ont été évaluées comme comparables, localement, aux données de références, pour des usages cartographiques. Mais ils ont également pointé les incohérences, l'hétérogénéité et les indécisions dans les descriptions. À ce moment-là, il paraissait qu'OSM ne pourrait corriger ses défauts qu'en s'imposant des contraintes semblables aux productions institutionnelles. Il est également apparu que des méthodes diversifiées d'évaluation de la base de données devaient être mobilisées en plus de la méthode de comparaison à une base de données de référence. En effet, les hétérogénéités spatiales et thématiques, ainsi que l'emprise mondiale de la base de données lui confèrent une originalité qui ne peut être rapportée à une autre base de données. L'étude des propriétés intrinsèques des données de la base de données OSM ont permis aux chercheurs d'exposer des tendances à la production de qualité en fonction du nombre de contributeurs, du nombre de versions des objets, de la qualité de objets voisins et surtout en fonction de la documentation originale du projet OSM. Ainsi, en développant des méthodes adaptées au mode de production du projet OSM, les chercheurs s'adaptent au caractère dynamique de la base de données. Comme la base de données est éditée en temps réel la fiabilité et la cohérence des informations produites ne peuvent être analysées à chaque publication semestrielle, par exemple. Les méthodes d'évaluation des données volontaires ont donc été adaptées à cette spécificité. Poursuivant cette recherche de diversification des méthodes d'évaluation, des scientifiques se sont tournés vers la question de la fiabilité. Si l'erreur n'est pas évitable, un producteur fiable a pour principal objectif de limiter la part d'erreurs dans ses publications.

OSM est apparue rapidement d'une hétérogénéité extrême, car le projet ne cessait de s'accroître en s'étendant géographiquement et en se diversifiant thématiquement. Cette

tendance entre en contradiction avec ce qu'on est en droit d'attendre d'une base de données crédible. On s'attend en effet à un système de production garantissant la réduction des erreurs dans les informations publiées. Il y a là un enjeu pour un bien commun. Les données ne peuvent être utilisables valablement, par exemple, lorsqu'un hôpital londonien ne présente pas les même attributs qu'un hôpital parisien. Ce deuxième enjeu, qui concerne la pratique de contribution, est associé au premier écart d'ordre méthodologique et académique d'absence de spécifications.

Enfin, OSM, progressivement devenu mondial, a été conçu sans usage mais peu de thématiques étaient envisagées dans les premières années. À la limite, le projet répondait à quelques usages individuels, les producteurs étant leurs propres utilisateurs et dialoguant aisément au sein d'un groupe restreint. Cela explique la *quasi* absence de spécification, aux débuts d'OSM.

OSM a su se perfectionner tout en conservant son projet initial de diversification. Une voie intermédiaire d'organisation s'est progressivement dessinée. Les contributeurs ne pouvaient plus en rester à leur pratique d'origine. Le développement d'OSM a imposé aux contributeurs l'introduction de spécifications, pour répondre à l'arrivée de nouveaux venus et à la complexification de leurs modèles. Mais cette introduction de spécification n'a pas été normalisatrice et n'a pas pris la forme de normes contraignantes. De nouvelles spécifications sont apparues, d'autres ont été précisées, mais elles sont demeurées indicatives. Les spécifications se précisant, les données existantes l'ont également été. Le processus de rectification s'est imposé comme la voie pour la poursuite du projet dans le temps. Ainsi, il est possible de produire des données de faibles qualités, c'est-à-dire répondant à peu ou à aucun usage, tout en assurant la correction et la validation des données par la suite.

Cette évolution est le résultat d'une maturation et d'une imprégnation de la culture du Libre. Elle a répondu et engendré de nouveaux besoins et a surtout été rendue possible par la présence active, au sein d'OSM, de professionnels et d'experts de domaines différents : informatique, information géographique, documentation, notamment.

La question de la fiabilité nous a, enfin, amener à reposer la question de l'autorité scientifique en matière d'information géographique. OSM ne prétend pas à s'imposer comme une autorité institutionnelle au sens où elle ne porte pas un message unique et organisé et au sens où elle ne joue pas le rôle de diffusion d'un message ou d'une vérité.

Tel n’est pas son objectif. Et l’une des caractéristiques présentes d’OSM est ne pas s’assigner un objectif clair et distinct. Cependant, OSM peut devenir une autorité dans le domaine de l’information géographique en proposant un nouveau mode de production. Elle pourrait alors être reconnue comme une source complémentaire d’information et même apparaître comme et unificatrice en matière d’interopérabilité.

Cette évolution permet de penser qu’on s’achemine vers un rapprochement et une complémentarité entre institutions et biens communs de la connaissance.

PARTIE I. Conclusion : OSM une source unificatrice ?

Dans cette partie, nous avons exposé un nouveau contexte de production de données géographiques, celui de l'information géographique libre. L'information géographique libre est une production formalisée dans un cadre non institutionnel mais qui sont réalisées selon les méthodes éprouvées du travail scientifique, que sont le processus de rectification, la reproductibilité des résultats et la vérifiabilité des informations. Ce type d'information, produit dans un contexte essentiellement numérique, présente des caractéristiques originales (une licence *copyleft*, une libre modélisation et l'accès à tous pour l'utilisation et la création), inspirées par une volonté politique de partage des connaissances et d'ouverture à destination des producteurs potentiellement non professionnels du domaine de la géographie. Ces motivations initiales ont ainsi mené à la création, en 2004, du projet OpenStreetMap. Le projet répond aux objectifs de partage des connaissances dont la conception est commune aux membres d'un collectif, constitué aujourd'hui de centaines de milliers de contributeurs actifs.

Le cas d'OSM est l'exemple principal de ce travail qui consiste à interroger les usages potentiels des informations géographiques libres. Dans le premier chapitre, nous avons décrit le changement de paradigme que connaît le domaine de la cartographie du fait l'informatisation et de l'apparition du procédé d'ouverture des codes informatiques et des données. Ce processus a mené à la généralisation des systèmes d'informations dont une grande partie des technologies sont déjà ouvertes et à l'émergence de l'*open data*. Après avoir détaillé les méthodes générales de formalisation de l'information géographique, nous avons exposé l'originalité de l'information géographique libre et du système de diffusion choisi par ses producteurs. Bien que les systèmes de production de l'information géographique libre s'inspirent de procédés plus traditionnels, ancrés dans les institutions, leur singularité nous a menée à poser la question de leur qualité, s'agissant des réalisations cartographiques. En tant que cartographe, nous attendons d'une source documentaire qu'elle soit produite selon un système de production garantissant la réduction des erreurs dans les informations publiées. Initialement, le projet OSM ne présentait que peu de garanties à un utilisateur extérieur au projet. Depuis, OSM a su se perfectionner, tout en conservant son projet initial de

diversification. Les contributeurs hybrident progressivement les méthodes traditionnelles de production d'informations géographiques avec les méthodes de production respectant les principes du Libre. Ils parviennent à imposer des spécifications à la fois pour répondre à l'arrivée de nouveaux venus et pour prendre en compte la complexification du projet. Opportunément, cette introduction de spécifications n'a pas été normalisatrice et le projet conserve sa force créatrice.

OSM, projet devenu mondial, est devenu un vecteur d'innovation. En effet, outre la base de données, dont le modèle horizontal est en constante mutation afin de présenter des informations au plus près des spécificités locales, OSM est à l'origine de la licence *copyleft* ODbL. Cette innovation spécifique s'inscrit dans le domaine juridique. La licence ODbL, spécifiquement conçue pour la protection des bases de données, s'inspire des travaux de pérennisation des projets de développement de logiciels relevant du Libre. La licence, conçue par des institutions internationales pour la production de connaissances ouvertes (Open Knowledge International) en réponse à la demande faite par la Fondation OSM, est une démonstration de la faisabilité d'une base de données ouvertes et protégées. Certaines institutions (comme nous l'avons vu, par exemple, la mairie de Paris ou la mairie de Toulouse) tendent à participer au mouvement de l'ouverture des connaissances et l'emploient ou s'en inspirent afin de répondre pleinement à leurs devoirs d'autorité publique (c'est ce qui a mené à la publication de la Licence Ouverte par Etalab, en 2012).

Les institutions ne se contentent pas de côtoyer OSM, elles développent progressivement des partenariats avec le projet. Nous avons exposé des formes de partenariats impliquant graduellement les parties. Une institution, volontaire pour participer au projet d'OSM, peut publier ses ressources documentaires dans le domaine public. Elle ouvre alors ses données à quiconque souhaite les exploiter. Autre possibilité : une publication engagée dans la production libre publie ses ressources sous une licence compatible, voire même sous la licence ODbL. Enfin, nous avons présenté des ententes spécifiques entre une autorité productrice et la Fondation OSM (comme la cession des droits par Bing de ses images satellites) ou avec une association OSM locale (avec les exemples de la ville d'Ottawa et de la Direction générale des Finances publiques en France). Les partenariats associant des institutions traditionnelles et le projet OSM portent ainsi deux objectifs :

- bénéficier respectivement des ressources de chacune des parties afin de construire un ensemble documentaire diversifié à portée géographique et

- participer aux mouvements politiques de l'ouverture des connaissances dont les effets attendus sont à la fois l'innovation et la construction d'une relation de confiance entre le citoyen et l'autorité.

Les partenariats avec OSM, pour une production d'information libre, ne sont pas cantonnés aux institutions traditionnelles. Ainsi, nous avons souligné la relation en construction entre OSM et Wikidata. Ce partenariat entre les deux projets est illustré par le dessin présenté ci-dessous (fig. 25). Il montre la construction d'un pont sémantique reliant OSM et Wikidata. Ce dessin est affiché à l'identique dans chacun des *wikis* des deux projets.

Figure 25 « OpenStreetMap-Wikidata Semantic Bridge », publié par Krauss en juillet 2018

Ce pont sémantique en construction symbolise les objectifs actuels des producteurs de connaissances. Grâce à ce pont, il s'agit d'abord de partager des moyens techniques, dans l'objectif de construire des bases de données de connaissances. Cela permet ensuite de déduire de nouvelles connaissances des grandes quantités d'informations détenues par chaque producteur. Ces informations se présentent sous deux types de forme, l'information libre telle que la base de données OSM et la production de folksonomies déstructurées. Nous présentons ces deux types d'informations produites en ligne et destinées à être intégrées dans ces bases de connaissances : l'information géographique libre, déjà abordée et les folksonomies déstructurées telle que la base Flickr de que nous qualifions ainsi afin de souligner la très faible formalisation par les

producteurs qui étiquettent les ressources sans se conformer à des normes. Les folksonomies sont des classifications en ligne, prises en charge par les internautes, et exercées par le moyen d'étiquettes affectées aux ressources. Entièrement libres dans leur écriture, les folksonomies sont produites par un public d'internautes, sans qu'il soit nécessairement conscient de son activité de classification. Si OSM peut relever de la folksonomie dans son sens large, la production de la base de données OSM présente un objectif d'organisation bien plus élaborées que celui de la base de données Flickr ou du balisage de page HTML par des particuliers. Ces formes de production d'informations sont rendues possibles par un nouveau contexte technique. Il permet une meilleure accessibilité aux méthodes et aux outils associés aux systèmes d'informations géographiques par le biais de la diffusion pédagogique de travaux de recherches et grâce au développement d'outils aux interfaces plus ergonomiques.

Alors que la production en ligne se simplifie et s'ouvre à des non experts de l'informatique, La production d'information en ligne se complexifie par une diversification des procédés de création d'informations. En s'ouvrant concrètement au grand public et aux interventions extérieures – et non plus seulement en théorie et légalement – les projets comme OSM, qui tendent à une structuration et à une reconnaissance de leur fiabilité, ont l'obligation de s'adapter aux attentes ainsi générées. Nous développerons donc, dans la seconde partie de la thèse, les enjeux qui accompagnent l'ouverture du système OSM à un public intéressé par le projet, sans nécessairement s'impliquer dans sa gouvernance. Les acteurs économiques sont, depuis les années 2010, des intervenants dynamiques au sein du projet OSM. Si le mouvement du Libre se concentre sur l'articulation entre la diffusion des connaissances et la commercialisation de services, dans quelle mesure l'action des entreprises sur la production des connaissances vient-elle perturber le système en construction d'OSM ?

PARTIE II : Cartographie industrielle et communs numériques

Jusqu'à présent, la thèse s'est attachée à considérer les producteurs d'informations géographiques, les méthodes de production et les informations elles-mêmes. Dans cette partie, nous changeons de focal et nous analysons plus particulièrement l'usage de l'information géographique libre dans un contexte industriel à travers les publications des acteurs-producteurs de données, mais également des acteurs qui diffusent ces informations et, parfois les commercialisent, après transformations.

Afin d'évaluer l'usage de l'information géographique libre dans un contexte industriel, nous nous appuyons sur des cas d'usages et plus particulièrement sur l'usage de la base de données OpenStreetMap au sein de l'entreprise Michelin Travel Partner. Au début de cette analyse nous nous demandons si l'intégration de l'information géographique libre dans un système de production industrielle le transforme radicalement, au point qu'il devienne un nouveau système de production. Et, respectivement nous posons l'hypothèse que l'action contributrice d'entreprises spécialisées dans le domaine de l'information géographique trouble le collectif déjà existant.

Pour répondre à ces questions nous décrivons, dans un premier temps, les usages commerciaux de ces informations. Nous distinguons les motivations qui mènent les industriels à emprunter des techniques et des modes de pensées issus du Libre. Dans un second temps, nous nous demandons si le système de production construit par les contributeurs à OSM réalise concrètement la théorie des Communs, qui trouverait ainsi dans ce mode de production une part de son application et de son actualisation.

Nous nous concentrons, dans le troisième chapitre, sur le processus de création cartographique menant à la commercialisation de carte et de services dérivés d'informations géographiques libres. Nous nous interrogeons sur la possibilité, dans le cadre d'une activité industrielle, de confectionner des cartes à partir d'une base de données libre (au sens de sa modélisation et dans sa production) et diffusée sous une licence ouverte. La carte ainsi produite devrait respecter la convention tacite entre son lecteur et son producteur, suivant laquelle la sélection éditoriale n'est pas dictée par une déficience d'informations. En outre, nous nous attachons à mettre en évidence les

inévitables conséquences graphiques de l'utilisation d'une telle source de données. La production de cartes papiers à portée éditoriale conçue à partir de base de données contributives présentes certaines difficultés. Celles-ci ne sont pas insurmontables et la preuve en a été faite par la publication de plusieurs cartes et guides par l'entreprise Michelin. Dans ce chapitre, nous décrivons d'abord l'offre commerciale basée sur de l'information géographique libre. Nous nous demandons, en effet, en quoi l'information géographique volontaire contributive libre et l'*open source* ont permis la poursuite ou l'émergence d'offres commerciales.

Cependant, c'est bien dans le cadre d'Internet que de nouveaux acteurs producteurs de créations cartographiques sont apparus. Le moyen qu'est Internet a changé le nombre de canaux d'informations permettant par sa gratuité relative, l'accès et surtout la publication de contenus exclusifs et spécifiques qui n'auraient pas trouvé un marché commercial en support papier. À partir des années 1990, de nouveaux acteurs émergent avec Internet et acquièrent une autorité méthodologique dont les propositions de standards tendent à s'internationaliser. Dans le quatrième chapitre, nous montrons que ce processus de standardisation au service d'échange d'informations est motivé par l'objectif de partage des connaissances entre pairs des chercheurs internationaux responsables de l'invention d'Internet qui, collectivement, développent les prémisses du réseau Internet. Puis, aux chercheurs s'ajoutent de nouveaux groupes sociaux qui prennent en charge le réseau, le diffusent et l'augmentent techniquement. De plus, un second processus transforme la production d'informations, également lié à l'internationalisation des échanges, c'est le mouvement du Libre qui est un courant d'idées au fondement de regroupements sociaux en ligne et parfois même à l'origine de communautés physiques. Nous nous intéressons à la partie fortement active en ligne de ces groupes qui sont formés de contributeurs. L'analyse montre que si ces groupes sont réels, constituant des associations internationales comme la Fondation OpenStreetMap, leurs organisations internes sont très lâches. Ils accordent une place importante à des acteurs industriels parfois moteurs dans le développement de productions contributives mais ils laissent aussi parfois un rôle proche du "passager clandestin" à des acteurs industriels inactifs dans la gouvernance169 du projet OSM. Nous observons les

169 Au sens d'Elinor Ostrom c'est-à-dire : « un système de règles et normes, définissant les droits (et obligations) attribués aux différents membres d'une communauté » (Oliver Weinstein dans Orsi et *al.*, 2017). Plusieurs définitions répondent au terme de gouvernance comme le rappelle Olivier Weinstein : « le terme de gouvernance est ancien. Il a été utilisé en France, au XIIIème siècle, dans le sens de mode de gouvernement […]. Il sera repris dans différents pays d'Europe, et notamment en Angleterre. Après être tombé en désuétude, il réapparaît et devient dans

procédés de prise de décisions entre contributeurs et entre le projet OSM (au travers de sa Fondation) et les acteurs extérieurs au projet. Dans ce chapitre, nous explorons les relations qui se constituent entre nouvelles institutions d'autorité (tels que le W3C et l'OGC), entreprises traditionnelles (tel qu'ESRI) et entreprises du numérique (tels que Google et Mapbox). Alors que le projet OSM est conçu pour la création d'une base de données créée par des bénévoles, dans quelle mesure la participation de ces acteurs dans l'écosystème d'OSM peut-elle le déséquilibrer ?

des écrits et les débats politiques à partir des années 1980 […] Le point de vue le plus général est que l'usage du terme de gouvernance, plutôt que celui de gouvernement, exprime un changement profond des fondements de l'ordre économique et politique, lié, comme nous allons le voir à la montée des idées néo-libérales et à l'affaiblissement des États. Cela exprimerait une volonté de dépolitisation des institutions au bénéfice d'une conception technocratique d'institutions « efficaces » » » (Oliver Weinstein dans Orsi et *al.*, 2017).

Chapitre 3 : Entre publications cartographiques commerciales et productions libres d'informations géographiques

L'essentiel des productions commerciales issues des bases de données contributives comme OpenStreetMap (OSM) est publié sur des supports numériques. Ces publications s'intègrent dans le Web des réseaux sociaux, des *tags*, des commentaires, grâce à la déclaration des utilisateurs de leur propre localisation ou de leurs recommandations de points d'intérêt. Le principal objectif d'OpenStreetMap est de mettre à disposition de cartographes professionnels ou amateurs autant de représentations d'objets de la réalité physiques et durables que possible. Le dessein affiché est de faciliter l'élaboration de cartes, au sens propre du terme, pour les cartographes professionnels et amateurs, à condition de respecter les conventions qui permettent la lisibilité et l'unité de l'image.

Le projet OSM et l'écosystème qui s'est développé autour de lui, offrent une diversité technique et documentaire inédite. Dès 2011, le service de cartographie de Michelin s'est interrogé sur l'opportunité d'utiliser la base de données produite et enrichie dans le cadre du projet OpenStreetMap pour produire une carte papier Michelin, notamment celle de la ville de Clermont-Ferrand au 1/12000ème. Trois arguments plaident en faveur de l'utilisation de cette source de données ouvertes. D'abord, le projet OSM s'est donné l'ensemble de l'Écoumène comme territoire. Sa base de données contient des informations suffisamment riches à échelles locales pour être utilisables en cartographie ou bien comme support d'accumulation de connaissances locales originales. Ensuite, les données sont accessibles sous la licence ODBL. Enfin, de nombreux outils d'édition (tels qu'ID, JOSM), des outils d'extraction (tels que Geofabrik, Qgis) et des outils de traitements (tels que Imposm3, PostgreSQL/PostGIS) lui sont dédiés ou adaptés. Ces outils et les activités mises en place dans le cadre du projet OSM visent ainsi à aider à l'exploitation de la base de données construite pour divers usages et à rendre son accès plus aisé à un public étendu, au dehors de la sphère professionnelle[170].

[170] Citons les outils osm2pgsql, QuickOSM intégrés à Qgis, le site Internet openstreetmap.org présentant la *Slippy Map* ou encore l'apparition d'entreprises de services de création de cartes en ligne personnalisées comme Mapbox ou JawgMap.

L'hétérogénéité sémantique et géométrique des données est d'une complexité inhérente au projet contributif. Un contributeur peut soit décider de lui-même de s'écarter du choix majoritaire d'écriture, soit y être contraint, faute d'information. L'utilisateur doit donc identifier les différentes écritures pouvant coexister dans la base de données pour une même information et faire le tri entre l'utile et le redondant. L'apport thématique est certes assuré par la liberté de contribution, mais les limites graphiques sont réelles à très petites échelles. La construction d'une chaîne de production cartographique industrielle peut représenter un apport dans le seul cas où ces deux paramètres sont mis en balance. En effet, il ne s'agit pas d'une faille du projet mais d'une incompatibilité entre l'échelle choisie par la communauté et la fonctionnalité d'accès aux petites échelles de toutes cartes numériques. Afin de répondre à ce besoin, des procédés sont inventés. Ainsi, une part considérable du temps de travail de cartographie est-elle consacrée à la recherche ou à l'élaboration de programmes informatiques en vue de favoriser, par exemple, l'extraction des données, le lissage de tracés ou l'ajout de nouveaux attributs. Les outils de cartographie et donc du cartographe changent alors fortement. L'ancrage dans un écosystème social et technique ouvert offre clairement un accès à de nouveaux outils sans délai de commercialisation. Le cartographe lui-même peut choisir de diffuser ses outils personnalisés.

Un des objectifs soulevés par cette thèse est de proposer des moyens d'exploiter cette accumulation de données pour produire une carte professionnelle à dix-neuf niveaux d'échelle, destinée à la mobilité des voyageurs et à la compréhension des territoires visités par les utilisateurs des cartes Michelin. Ces nouvelles formes de productions et ces nouvelles réalisations reposent sur des API Web de cartographie. Une API est un moyen technique de développement, diffusé sous la forme d'une bibliothèque de fonctions ou de méthodes ou sous une forme plus intégrée en service Web. Une API sert à faciliter l'intégration d'un produit dans d'autres produits. Des API peuvent être développées pour tout usage, du service de paiement en ligne à l'intégration aisée de la carte Google Maps (une ligne de code dans un document HTML) dans une page Web ou une application. Une API Web peut donc être le principal service commercialisé par une entreprise. L'entreprise vend alors son expertise technologique sous la forme de ce service. L'entreprise développe et diffuse son API également dans le but de rendre son service intégrable à d'autres services

numériques par des développeurs et de constituer ce que l'on nomme « un écosystème de services ».

Ces API se présentent, en cartographie, sous la forme de bibliothèques de développement (Leaflet, Open Layers, Mapbox, Google Maps) qui sont des moyens de faciliter l'affichage de cartes et qui permettent l'interactivité avec des objets de la carte ou simplement le *mashup cartographique*[171]. L'utilisation à visée cartographique d'API Web s'est manifestée par la diffusion rapide de la carte Google Maps sur tout type de sites Web, destiné à la localisation de points d'intérêts ajoutés par l'utilisateur non développeur. Par exemple, le service Uber exploite l'API Web et la bibliothèque de développement de Google Maps. Uber bénéficie ainsi de la technologie de géolocalisation de Google et externalise cette compétence centrale pour son propre service d'échange de positions entre les clients et les conducteurs. L'utilisation du projet OpenStreetMap ne se réduit pas à la publication d'un fond de carte au service de *mashups*.

Le projet OSM est une des plus importantes sources de données ouvertes menant à une diversification des usages, notamment en ligne. En 2019, on recense des cartographies mondiales généralistes (openstreetmap.org, les fonds de carte des entreprises Mapbox et Jawg), des cartes relatives à la pratique de sports (comme le cyclisme[172], le ski[173] et la randonnée[174]), des cartes de transport (en commun et pour l'emprise mondial[175] et locaux[176] et maritime[177]) et des cartes présentant des services et des infrastructures publics (accessibilité aux lieux en fauteuil roulant : https://wheelmap.org). La liberté de contribution offerte par le contexte de production OSM a mené à des innovations cartographiques comme la modélisation de l'intérieur de bâtiments publics (https://openlevelup.net) et l'exploitation des nouvelles méthodes de description d'objets 3D (https://demo.f4map.com). Cette émulation, rendue possible par l'accès à

[171] Forme de représentation de données spatialisées en superposition d'un fond de carte. Le *mashup* est rendu accessible à un très grand public par la construction de services à cet effet comme celui de Google Maps ou de uMap (utilisant des fonds cartes issus d'OSM, https://umap.openstreetmap.fr/fr/)
[172] http://www.opencyclemap.org/
[173] http://www.opensnowmap.org/
[174] http://hikebikemap.org/
[175] style transport du site http://www.openstreetmap.org
[176] http://www.öpnvkarte.de
[177] http://openseamap.org

une base de données mondiale, ouverte à la contribution, est un terrain d'expérimentation infini pour la recherche.

Deux types d'expérimentation sont à distinguer.

D'une part, OSM est utilisé comme une source de données permettant d'enrichir des recherches. C'est notamment le cas dans le domaine de la reconnaissance de formes urbaines (Liu et Long, 2016 ; Liu et *al.*, 2017 ; Tao et *al.*, 2017) ou dans celui de l'occupation des sols (Estima et Painho, 2013, Estima et Painho, 2015 ; Johnson et Iizuka, 2016). OSM fonctionne alors comme une source de données relatives à l'environnement pour des sujets autour de questions de santé comme la pollution (Mueller et *al.*, 2018), l'accès aux services de santé (Mooney et Corcoran, 2013), la prise en compte du bâti pour le déploiement d'infrastructures de diffusion d'ondes radios (Nuckelt et *al.*, 2013) ou l'évaluation de potentiels de mobilité entre espaces frontaliers européens (Guérois et *al.*, 2016, Marquet, 2016).

D'autre part, le contexte Libre de la base de données OSM offre aux chercheurs l'accès à une grande quantité de données dont l'usage est réglementé – non pour le limiter mais pour l'encourager – à l'inverse d'un service comme celui de Google Maps, par exemple. Ainsi, il est possible d'utiliser OSM comme une source de données de géocodage. Dans le contexte de la recherche et du respect de l'éthique, les données à caractère personnel ne peuvent être localisées à l'aide de ressources propriétaires. En effet, les fournisseurs de ces ressources peuvent s'approprier les données initiales et le résultat de la transformation qu'elles permettent. En outre, la base OSM constitue un moyen d'expérimentation d'implémentation d'algorithmes pour la navigation routière d'après des données réelles et non théoriques (Hentschel et Wagner, 2010, Roussel et al, 2015) et pour l'extension des modèles de description d'objets en 3D (Goetz, 2013 ; Boeters, 2015). En dépit d'un constat d'hétérogénéité, les données libres sont une opportunité pour la standardisation des procédés testés, qui doivent opérer aussi bien sur des données prévues à cet effet que sur des données hétérogènes et changeantes (Over et al, 2010).

Comme nous l'avons vu dans le deuxième chapitre la confiance en une source d'informations est relative à la qualité intrinsèque des informations produites, d'une part, et au caractère opérationnel de ces informations, d'autre part. L'ensemble de ces informations est hétérogène. Elles demandent donc l'application de procédures

d'homogénéisation pour les représenter graphiquement, selon les mêmes signes et les mêmes codes. La représentation graphique d'informations géographiques par des figurés élémentaires se pratique, classiquement, selon un ensemble de conventions partagé entre cartographes. Ces conventions sont connues des utilisateurs de cartes grâce à la légende de la carte ou aux habitudes de lecture.

De façon générale, une source d'informations est exploitée dans le but d'en déduire des connaissances puis, dans notre cas, de les représenter sous forme de carte. Nous désignons ce processus de transformation du matériau initial (un ensemble d'informations géographiques) jusqu'à la commercialisation du produit fini (la carte) par l'expression « chaîne de production ». Dans ce chapitre nous expliquons l'intégration de la source de données OSM dans le processus de création cartographique de Michelin. Une chaîne de production cartographique se construit en prenant en compte les contraintes imposées par le contexte de production. De manière générale, les contraintes sont d'abord économiques. La chaîne de production doit répondre aux objectifs sans dépasser un certain budget investi. Ce budget se répartit en investissements matériels et en temps de travail. Dans le cas de la production de cartes, l'objectif est la publication finale du produit respectant une ligne éditoriale et les qualités graphiques que sont la lisibilité et l'harmonie graphique. Les contraintes techniques sont inhérentes à la définition d'une chaîne de production. Il s'agit de trouver un équilibre entre investissement dans des outils chers mais fonctionnels dans l'immédiat et simples d'utilisation, et l'investissement dans des techniques spécialisés aux besoins métiers, nécessitant la formation des employés et des compétences de fabrication d'outils interne à l'entreprise. Dans le cas d'OSM, les chaînes de production que nous avons conçues n'ont pas été exemptes de ces contraintes. L'écosystème OSM présente certaines techniques originales, dont la contribution, la description par la folksonomie[178] mais également les outils de lecture et de manipulation des données. Ces aspects techniques ont dû être acquis par les utilisateurs des données OSM au sein de Michelin. En outre, le cahier des charges définissant les objectifs du projet exigeait que les nouvelles chaînes de production mobilisent des techniques déjà ancrées chez

[178] Rappelons que le terme folksonomie désigne la pratique de classification d'informations (photographies, pages web, signets) par les utilisateurs d'Internet qui marquent les ressources disponibles sur le réseau par des étiquettes (ou *tags*). La folksonomie est définie par les auteurs, s'appuyant sur Wal T.V. in « Folksonomy Definition and Wikipedia » (2005) comme « le résultat de l'activité d'association libre d'étiquettes (*tags*) à des données ou objets (tout ce qui est identifié par une URL) dans le but personnel de les retrouver ultérieurement » (Hombiat et *al.*, 2015, p.106).

Michelin. La chaîne de production cartographique est donc envisagée comme une méthode composite, qui allie adaptation aux méthodes traditionnelles du métier et appropriation de techniques extérieures.

L'entreprise centenaire a défini progressivement des codes graphiques désormais précis et relativement fixes, au service de la mobilité et du tourisme. Les cartes et les plans doivent apporter rapidement les informations nécessaires au déplacement de l'utilisateur en direction de lieux ou de points d'intérêts mis en évidence par la ligne éditoriale. Dans ce contexte de création, les qualités de forme indispensables à l'expression et à la réception de l'information géographique sont des éléments centraux du travail du pôle cartographique de Michelin. Le soin esthétique constitue un temps de production considérable afin d'assurer la lisibilité de la carte tout en présentant une forte densité d'informations. En exploitant les données OSM, nous nous trouvons confrontés en plus à un défi d'ordre graphique. L'hétérogénéité de renseignement des données, peut conduire à des territoires disparates en termes de densité de données. Un véritable défi doit être relevé lorsque ces écarts de densité ne sont pas le reflet d'écarts de densité dans la réalité mais le reflet des écarts de contribution. Par ailleurs, ces données sont peu hiérarchisées. L'hétérogénéité comme la faible hiérarchisation des données empêchent alors la constitution d'une chaîne de production cartographique très largement automatisée. En effet, une chaîne construite dans le but de dé-densifier les données risquerait d'appauvrir davantage les territoires peu couverts. En outre, les objets étant faiblement hiérarchisés, l'automatisation de dé-densification ne peut reposer seulement sur les *tags* parfois trop peu discriminants. La création d'une nouvelle chaîne de production chez Michelin n'a ainsi pas constitué un nouveau pas vers l'automatisation de la production cartographique mais un retour à la participation des cartographes dans l'agencement des objets graphiques et dans le choix éditorial afin d'intégrer de la hiérarchie dans les objets en fonction de la visée de la carte ou du plan. Mais il s'agit sans doute d'un investissement nécessaire pour conserver la qualité des produits réalisés par l'entreprise.

Les conséquences du changement de sources de données sur la définition d'une chaîne de production industrielle sont méthodologiques. D'abord, l'information géographique libre pose la question de son intégration dans l'organisation linéaire de la création d'une carte. La cartographie, telle qu'elle est pratiquée et enseignée à l'université, suit la chaîne suivante : l'acquisition d'informations, la manipulation d'informations et le

dessin de la synthèse géographique de ces informations. La définition de l'image se situe souvent avant la phase d'acquisition, parfois entre acquisition et manipulation. Au contraire, l'utilisation des données libres implique le plus souvent la pratique de la contribution. Une chaîne de production initialement construite sur l'exploitation d'une source de données considérée comme stable mais également non modifiable, ne peut intégrer une source de données libres. Nous présentons, dans ce chapitre, les différences radicales entre une chaîne de production linéaire construite à partir de données propriétaires et une chaîne de production itérative permise par les principes du Libre. En d'autres termes, la cartographie, qu'elle soit sur un support numérique ou sur un support papier, intègre le processus de création de données au processus de création de documents cartographiques. Le producteur de cartes acquiert le savoir-faire de la collecte. Dans le cas de Michelin, il s'agit d'un retour à ce savoir-faire qui a fait l'originalité des premières cartes Michelin et à son balisage systématique des routes.

Décrire l'offre commerciale, ce que nous proposons de faire dans ce chapitre, permet d'exposer les utilisations effectives et de constater certains écarts avec les productions réalisées à partir des données propriétaires, qu'elles soient contributives ou non. Ces écarts suscitent parfois une insatisfaction des utilisateurs. Cependant, un produit dont la publicité met en valeur l'utilisation de la donnée libre, à l'aide d'outils libres, dans le respect des licences et publié par une entreprise qui supporte ces contextes de production, bénéficiera d'un *a priori* favorable. Une partie du public est capable de se montrer bienveillant à l'égard d'un produit issu du Libre, opposé dans son esprit aux produits diffusés par des entreprises traditionnelles et aux systèmes de production reposant sur des technologies fermées. Cependant, cette clientèle attentive au contexte de production demeure encore limitée. Nous verrons qu'elle est également exigeante quant au respect des licences et de la qualité des contributions aux projets libres. Opportunément, la motivation des entreprises à se tourner vers les données contributives libres n'est pas fondée seulement sur ce public soucieux des technologies libres, encore bien trop faible. Notre thèse s'attache à décrire tous les enjeux qui découlent d'une production cartographique industrielle, issue de données libres et fabriquée avec les outils du Libre.

L'objectif n'est pas de dévaloriser les productions basées sur des données fermées. Depuis le début des années 2000, celles-ci demeurent des références en tant que productions exploitées pour des cartographies très grand public. Cependant, recourir

aux données contributives permet au producteur de diffuser des produits enrichis d'informations alternatives et de publier des cartes complémentaires aux productions traditionnelles. Chacun de ces produits repose finalement sur des choix de système de production si différents, qu'ils ne peuvent s'opposer mais sont à considérer comme complémentaires. On ne peut cependant pas exclure que les attentes des utilisateurs envers l'ensemble des produits cartographiques fabriqués se voient modifiées par les données libres et, plus généralement, par la pensée du Libre.

Le système de la production par le Libre s'est profondément transformé depuis ses premières définitions du début des années 1980. Ce changement tient essentiellement à sa diffusion mondiale, dans tous les domaines concernés, par l'expansion des moyens numériques. Le développement et la création de ressources documentaires, qu'il s'agisse d'informations géographiques, de textes encyclopédiques ou de photographies ont, un temps, été largement bénévoles et pratiqués en dehors du cadre professionnel officiel. Dès 2011, Lin (2011) montre la diversité des facteurs d'implication en mettant en évidence par des entretiens semi-directifs des types de contributeurs dont l'activité contributive peut être rémunératrice au sein d'une entreprise, d'un gouvernement, d'une ONG ou bénévole (en s'intégrant à une structure ou non). Ce constat précoce se confirme avec l'apparition d'entreprises dont le système économique se fonde sur la disponibilité de la base de données OpenStreetMap. Les entreprises comme Mapillary et Mapbox, construisent deux systèmes économiques qui nécessitent une amélioration constante et rapide du projet OSM en général. Si elles tentent de bâtir des communautés contributrices annexes à OSM, autour de leur propre projet, elles intègrent au sein de leur société des services de contribution[179]. En outre, comme le soulignent Quinn et MacEachren (2018), intéressée par ces nouvelles dépendances économiques entre les contributeurs – qu'ils soient bénévoles ou non – et des entreprises, la Fondation OSM a émis en novembre 2018 des directives à destination des acteurs économiques qui organisent la contribution dans leurs propres intérêts[180]. Le préambule rappelle que la communauté OSM est, à l'origine, constituée par des contributeurs bénévoles.

[179] Dans le cas de Mapillary, les employés contributeurs participent davantage au projet Mapillary qu'à celui d'OSM, cependant l'entreprise publie également des outils d'aide à la contribution à OSM. Dans le cas de Mapbox, une équipe est concentrée sur la contribution à la base de données OSM. Mapbox présente ainsi sur sa page du *wiki* OSM les pseudonymes de contributeurs des quatre membres de cette équipe. Cette page documente explicitement les activités de l'équipe dans une optique de transparence avec la communauté afin d'assurer qu'aucune manipulation malveillante ou de publicité n'est pratiquée par l'équipe.

[180] La pratique visée est désignée par l'expression «*organised editing*» (*OSM Foundation*, 2018).

Cependant, la popularisation du projet a généré de nouvelles activités économiques délicates à définir. Les acteurs économiques emploient désormais des contributeurs salariés et encouragent même des campagnes de contribution non rétribuées. Et cela, d'une manière jugée parfois dirigiste par la Fondation. Ne pouvant nier que cette situation est de plus en plus fréquente, la Fondation a produit des directives à destination de ces acteurs. Il s'agit d'assurer le respect de pratiques approuvées par l'ensemble de la communauté à l'issue d'un consensus (OSM Foundation, 2018). La Fondation propose d'ailleurs une conceptualisation intéressante de ces contributions guidées par un intérêt économique. Selon elle, il s'agit d'éditions substantielles et organisées, dont les règles établies s'appliquent à plus d'un contributeur.

Afin d'établir une relation transparente avec la communauté mais également d'assurer une bonne communication avec cette communauté, la Fondation requiert de l'acteur économique qu'il documente son activité sur OSM dans le *wiki* du projet. La création de cette page du *wiki* est présentée comme une première étape vers l'échange avec la communauté. Il est attendu que l'échange se poursuive de manière ouverte, en ligne, par le moyen des forums et des listes de discussion classiquement utilisés par la communauté. Enfin, la Fondation rappelle qu'en cas de comportement non conforme aux règles établies par la communauté, qui s'appliquent individuellement à chaque contributeur, elle dispose de droits lui permettant de supprimer des contributions ou de bannir des comptes de contributeurs.

En 2014, la définition, par la Fondation, de bonnes pratiques à destination des acteurs économiques contributeurs à OSM commence à être débattue au sein de la communauté au travers de discussions en ligne, de propositions de textes par des membres de la communauté et par une enquête menée par la Fondation. L'intérêt apporté à cette question par la Fondation s'explique par la part importante prise par ces acteurs économiques dans la croissance du projet (comme cela est indiqué dans les directives citées)[181] mais également par la nécessité pour les acteurs économiques – actuellement en place – de garantir le développement d'OSM devenu le socle de leur système de production.

[181] «*OpenStreetMap is powered by its community. While originally supported by individuals, the continuing growth and popularity of OSM have also spawned organised mapping efforts by companies employing mapping teams and unpaid groups like school classes that are directed to work on OSM.*», *Organised Editing Guidelines* (2019), consulté en ligne en avril 2019.

Dans ce chapitre, nous explorons les pratiques et les relations commerciales qui existent en périphérie et au sein du projet OSM. Grâce à notre pratique de la cartographie au sein de l'entreprise Michelin, nous avons observé la force facilitatrice que certaines entreprises comme Mapbox ont conçue, afin de produire des cartes en ligne. En parallèle, des outils ont été conçus dans des contextes de recherches publiques ou privées, puis leurs codes ont pu être ouverts. Indispensables aux systèmes de production que nous avons mis en place chez Michelin, ils ont été utilisés indépendamment de tout service commercial (1. Systèmes de production cartographique). Ce sont ces relations soit de dépendance, soit de partage qui nous ont menées à explorer plus précisément la construction de l'écosystème technique et commercial d'OSM. Dès le début des années 2010, des acteurs économiques comme Google et Microsoft se sont connectés à cet écosystème. Au milieu des années 2010, de nouveaux acteurs ont émergé (fig. 26) ; ils ont essayé de construire des projets lucratifs au fondement desquels se trouveraient des communautés (2. Services et produits cartographiques dérivés de l'information géographique libre). Les relations de partage que nous évoquons peuvent se matérialiser au travers de projets favorisant l'interopérabilité et notamment entre bases de données institutionnelles. Les perspectives de poursuite du partage des connaissances, géographiques dans notre exemple, peuvent mener à une interopérabilité entre les bases de données institutionnelles et les bases de données folksonomiques (3. Exploitation des services et expérimentation sur l'interopérabilité de sources ouvertes pour la cartographie touristique)

Relations entre acteurs du contributif et utilisateurs du contributif dans le contexte industriel

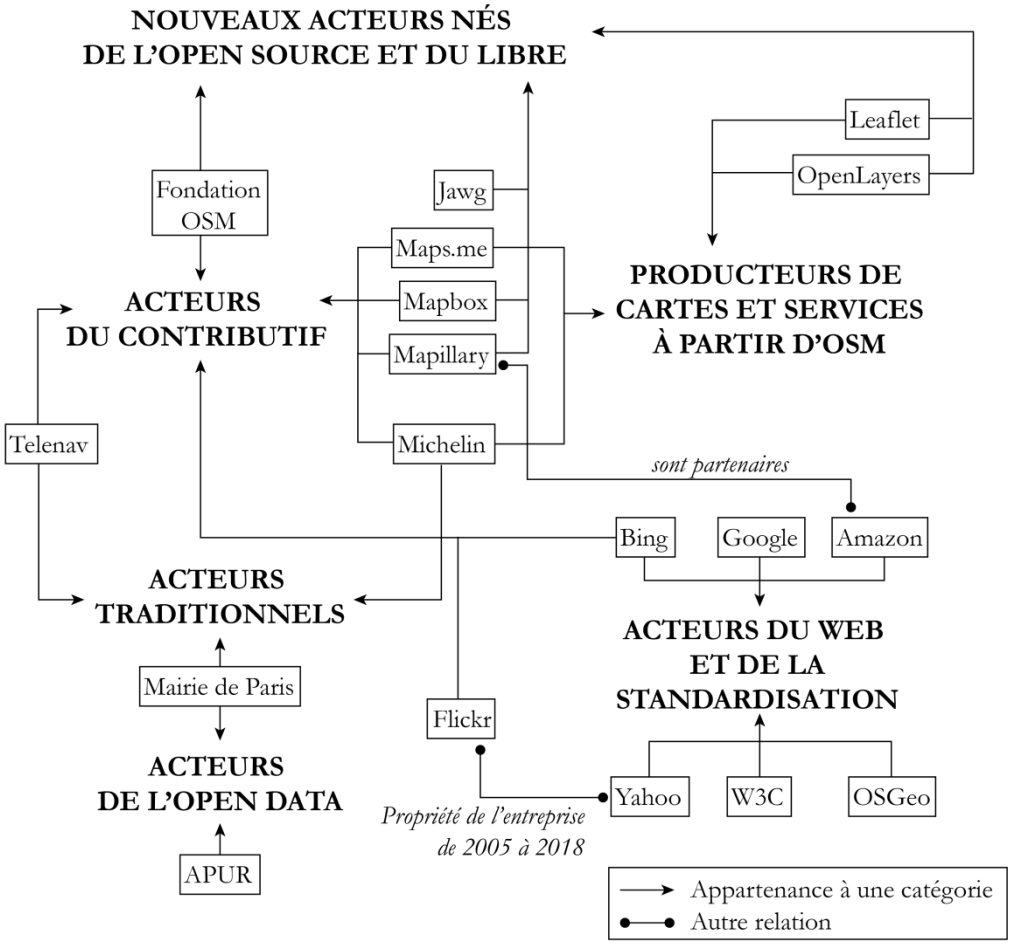

Figure 26 Acteurs et relations entre acteurs examinés dans ce chapitre (Flora Hayat, 2019)

3.1. Les systèmes de production cartographique

Le contexte de production de la base de données OpenStreetMap bouleverse la pratique de la cartographie par sa nouveauté. On s'appuiera, pour détailler les conséquences, sur des exemples de chaînes de production mises en place chez Michelin durant la thèse. La première chaîne de production a permis de tester la faisabilité de production d'une carte sur support Web et l'autre sur support papier. La production de cartes chez Michelin est guidée par une ligne éditoriale générale : aider l'utilisateur à la mobilité. Les choix graphiques et thématiques destinés aux cartes et aux plans doivent y répondre. La seconde chaîne de production, pour support papier est encore davantage guidée par des choix éditoriaux. En effet, la sémiologie et la sélection éditoriale des cartes Web sont uniformisées pour l'ensemble de l'Écoumène, ou au moins à l'échelle des pays, afin de prendre en compte les chartes graphiques locales pour la lecture du réseau routier par exemple (ViaMichelin présente une représentation du réseau spécifique au Royaume-Uni et en Allemagne, et Google Maps fait de même au Japon par exemple). Les plans de ville et les cartes régionales présentent certes des grandes lignes éditoriales par collection mais elles sont également largement adaptées aux territoires cartographiés.

Deux grandes différences distinguent la méthode de cartographie pour le Web de la cartographie pour une publication papier : l'emprise et l'échelle. En effet, la carte en ligne que nous avons cherché à produire devait être globale et lisible sur une large plage d'échelles. La cartographie pour le Web, telle qu'elle est présentée par ViaMichelin, permet de publier un ensemble de cartes à partir des échelles locales (bâtiments, noms et adresses) jusqu'à une échelle mondiale tout en proposant un service de calcul d'itinéraires. Mais cette quantité de données ne peut pas être représentée aussi clairement qu'elles pourraient l'être sur une carte papier. En effet, le processus d'édition manuel du dessin est une partie intégrante du processus de production d'une carte papier et ne peut pas être appliqué de la même manière pour les échelles locales d'un service de cartographique mondiale. Néanmoins, la fonctionnalité de zoom est le facteur qui compense l'écart en densité d'informations aux mêmes échelles entre cartographie papier et cartographie en ligne. Une carte papier doit suffire, à elle seule, à transmettre le message. Lors de la consultation d'une carte Web, si une information n'est pas claire

sur une échelle moyenne, l'utilisateur peut facilement changer d'échelle et obtenir les informations qu'il recherche.

Ces deux chaînes, conçues ici en tant que prototypes, ont finalement mené à des publications de cartes réalisées entièrement avec la base de données OSM dans des Guides Verts, des Guides Rouges et des cartes régionales. Nos expérimentations ont été réalisées selon des échelles et des territoires pour lesquels une étude marketing concluait à un besoin de la part de la clientèle habituelle de l'entreprise ou comme un moyen de capter de nouveaux lecteurs. Comme nous nous situons dans un processus de recherche et de production quasi simultanée, les étapes d'évaluation de qualité, au préalable du déclenchement d'une production, ont été réalisées de manières diverses. Un prototype rudimentaire est systématiquement réalisé, selon l'échelle et le territoire commandé. Ensuite, deux possibilités se présentent. Soit un cartographe et un documentaliste, experts du territoire à cartographier, évaluent la réalisation selon leurs méthodes traditionnelles, soit les outils, décrits dans le deuxième chapitre, sont mobilisés. Au terme des trois ans, un assortiment de ces méthodes a été sélectionné par les cartographes afin d'évaluer la faisabilité du projet. Malgré ces procédures méthodiques, des outils d'évaluation spécialisés, spécifiques aux besoins de la cartographie, demeurent à développer chez Michelin. C'est d'autant plus nécessaire que contribuer, dans le but de bénéficier de données actualisées et conformes aux originalités du territoire, constitue désormais une part fondamentale du processus de création tel qu'il a été conçu chez Michelin, par l'ensemble de l'équipe de la production cartographique.

3.1.1. Les chaînes de production

3.1.1.1. Une base de données cartographique issue d'une base de données folksonomiques

Comme première étape de la traduction des données OSM vers une carte, nous traduisons la base de données horizontale au format XML, qui est une copie de la base de données OSM en un modèle aussi simple et complet que possible. Ce modèle sert de bases à de futures modifications vers une base de données cartographique.

La première étape de construction d'une chaîne de production cartographique à partir d'OSM consiste à choisir l'outil qui va permettre d'enregistrer les données sources en un format utilisable pour des traitements automatiques par d'autres outils de réalisation cartographiques. Trois outils ont été identifiés :

- ***Osmosis*** est un programme en ligne de commandes permettant de manipuler et de traiter des données OSM. Il est utilisé pour traiter les extractions complètes de la base monde (volumineuse), pour diviser des fichiers .osm en éléments de plus petite taille (par zones géographiques par exemple) et pour mettre en place un différentiel afin d'en assurer la mise à jour. Osmosis est l'outil le plus complet permettant de manipuler les fichiers .*osm*. Il est néanmoins le moins adapté aux choix thématiques nécessaires à la construction d'une base de données pour la cartographie. Cet outil est le plus souvent mobilisé en amont du chargement des données, pour filtrer et ne retenir que certains *tags*, par exemple.

- ***Osm2pgsql*** est un programme en ligne de commandes permettant de charger des données OSM extraites de la base de données source dans une base de données PostGIS. Les données de type *data.osm.pbf* sont chargées par défaut en quatre tables de géométries (*planet_osm_point*, *planet_osm_line*, *planet_osm_roads* et *planet_osm_polygon*) et les *tags* déclarés dans le fichier de configuration sont transformés en colonnes. Osm2pgsql prévoit également le chargement de *tags* en une colonne de type *hstore*. Le type d'écriture d'une colonne d'une table en *hstore* est une fonctionnalité du système de gestion de base de données PostgreSQL. Ce moyen permet de stocker un ensemble de couples clé / valeur comme une seule valeur. Ce formalisme permet d'éviter un trop grand nombre de colonnes (clés) rarement renseignées. Dans le cas d'OSM, de nombreuses clés peuvent présenter un intérêt dans une table ; ce qui rend la table difficilement manipulable. C'est donc dans le cas de données semi-structurées que cette écriture en une colonne d'un ensemble de clé / valeur est utile.

- ***Imposm3*** est un programme en ligne de commandes qui permet de concevoir des schémas de base de données personnalisées. L'ensemble des tables et de leurs colonnes sont paramétrables, permettant de distinguer, par exemple, les tables par géométries et par *tags* principaux, tout en complétant ces *tags* principaux de leurs *tags* secondaires enregistrés dans des champs. Le programme prévoit une fonction

permettant d'unifier les valeurs (par exemple, les valeurs booléennes 1, *true* et *yes* deviennent toutes TRUE). Une fonction de généralisation est également prévue. Une colonne *hstore* peut être ajoutée à toutes les tables afin de stocker chaque *tag* dans un seul champ, évitant ainsi de perdre des *tags* mineurs mais utiles localement.

Au final, nous retenons, l'outil Imposm3 (pris en charge par Omniscale) lors de l'importation de données OSM dans une base de données spatiale. L'outil permet des choix de chargement personnalisés en termes de construction de modèle de données et se montre plus performant qu'osm2pgsql de manière générale. Un fichier de spécification (.yml) sert à définir un modèle personnalisé d'extraction des données OSM (fig. 27).

```yaml
tables:                 # Début de la liste des tables
 roads :                # Table des routes
  fields:               # Les champs de la table
  - name: osm_id        # Champ portant l'identifiant unique
    type: id
  - name: geometry      # Champ portant la description de la géométrie de l'objet
    type: geometry
  - name: valeur        # Champ portant la clé principale décidé dans le "mapping"
    type: mapping_value
  - key: name           # Champ portant le texte (string) enregistré dans la valeur de la clé name si
                          la valeur existe
    name: nom
    type: string
  - key: bridge
    name: pont
    type: boolint       # La valeur associée à la clé bridge est censée être contribuée sous une forme
                         booléenne (0, 1 ou 'yes', 'no'), Imposm3 interprète ces valeurs
  - name: tags          # Champ qui contient toutes les clés décrivant l'objet
    type: hstore_tags
  filters:
   exclude_tags:     # On ne souhaite pas de routes (highway) qui decrivent une zone (comme des places)
   - -area
     -'yes'
  mapping:
   highway:
   -_any_               # On retient toutes les valeurs possibles associées à la clé highway
   type : linestring
```

*Figure 27 - Exemple de déclaration des tables avec Imposm3, ici une table (*roads*), ses champs (*osm_id, geometry, valeur, name, tags*) et sa géométrie (*linestring*), en marron les commentaires de l'auteure explicitant les paramètres écrits (Hayat, 2019)*

Nous disposons, à ce stade, d'une base de données généraliste qui permet la constitution d'une chaîne de cartographie pour le support en tuiles vectorielles comme pour le support papier. La contrainte des performances des outils de visualisation des cartes numériques implique que les tuiles vectorielles contiennent uniquement les données nécessaires pour correspondre aux attentes éditoriales. À l'inverse, une base de données

à destination de productions papiers doit présenter une plus grande diversité thématique afin de répondre aux attentes plus spécifiques des chartes graphiques et éditoriales des cartes Michelin. La sélection et la hiérarchisation de l'information sont des méthodes de cartographie traditionnelle mais, dans le cas d'OSM, elles sont à pratiquer avec davantage d'amplitude. Le cartographe recherche un compromis entre une figuration didactique du territoire et une exploitation fine des informations présentes dans la base OSM.

3.1.1.2. Les tuiles vectorielles

La technologie vectorielle présente plusieurs intérêts qui la distinguent de la technologie raster. Tout d'abord, elle bouleverse les rôles de serveur et de client. La carte est affichée par une bibliothèque chargée de la mise en forme côté client, la bibliothèque Mapbox GL JS (fig. 28, n°6), dans notre exemple. Les tâches de visualisation sont donc réalisées par l'appareil de l'utilisateur. Lorsque l'utilisateur navigue dans la carte (fig. 28, n°1 et n°8), l'application de cartographie demande au serveur les tuiles vectorielles (fig. 28, n°2). Elles ne sont alors que des jeux de données sans représentation. Les tuiles transférées couvrent l'emprise de visualisation et l'échelle choisie par l'utilisateur (fig. 28, n°3 et n°4). Chaque tuile contient une sélection de données géographiques et leurs attributs. Elle peut être construite, selon les paramètres du cartographe, pour être affichée en fonction d'une plage d'échelle, en continu. Les données vectorielles sont stockées dans le cache du navigateur, côté client (fig. 28, n°5).

L'atout majeur de la technologie des tuiles vectorielles pour la cartographie est la possibilité pour l'utilisateur de changer de charte graphique, décrite sous la forme d'une feuille de style, sans nouvel appel de données. Grâce à cette organisation qui place l'étape de mise en forme en bout de chaîne, c'est l'appareil de l'utilisateur qui met en forme la carte (fig. 28, n°6). Tant que le cartographe prévoit des chartes graphiques qui s'appliquent à la même structure de base de données, le changement n'exigera pas un nouveau téléchargement et la bibliothèque d'affichage de la carte remettra en forme l'image.

Le stockage des tuiles dans le cache de l'appareil de l'utilisateur permet, une fois le délai de chargement passé, une navigation fluide et rapide dans la carte. Cette navigation fluide repose également sur la composition des tuiles vectorielles, qui ne

sont pas des images fixes mais des jeux de données évolutifs. L'action du zoom par l'utilisateur déclenche la mise à l'échelle et le réglage de position des icônes et des étiquettes, sans nécessiter forcément une nouvelle requête au serveur si la plage d'échelle concernée est déjà stockée dans le cache de l'appareil de l'utilisateur (fig. 28, n°7).

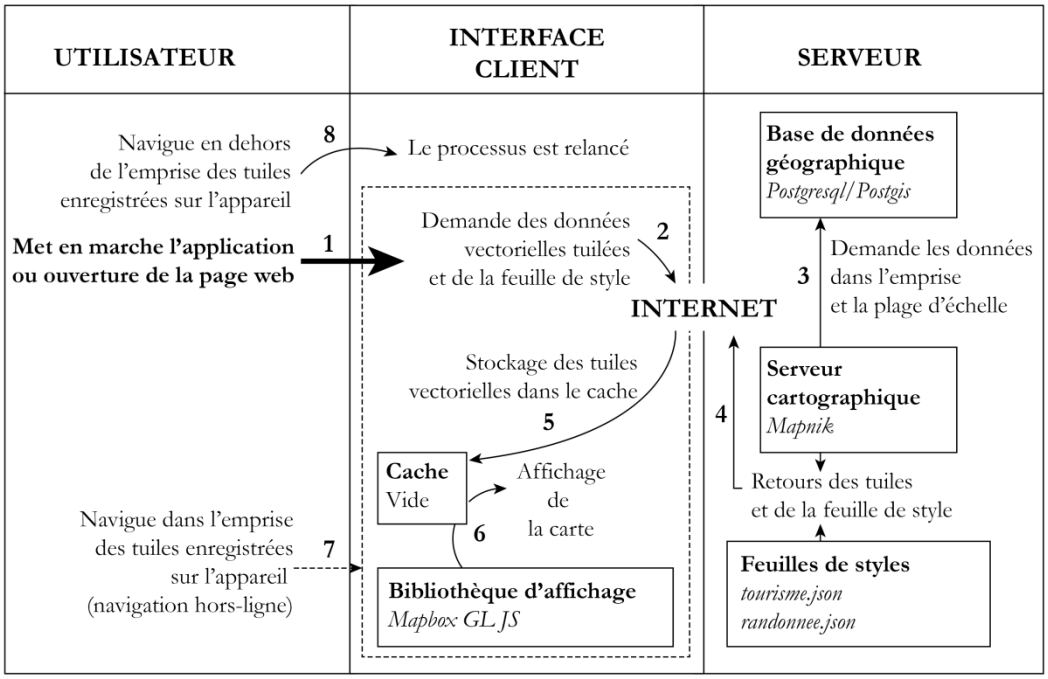

Figure 28 Requêtes auprès du serveur par le client. (Hayat, 2019)

3.1.1.3. Les tuiles vectorielles et les modèles de données

Afin de progresser dans la construction du prototype d'une carte Web Michelin, une chaîne de production cartographique complète est mise en place. Deux outils (Utilery et Tilelive) ont été utilisés pour produire des tuiles vectorielles. Utilery (développé par Yohan Boniface) produit des tuiles vectorielles qui s'intègrent dans la technologie développée par Mapbox. Il est ainsi possible d'utiliser la bibliothèque d'affichage cartographique proposée par l'entreprise. Dans le fichier (.yml), il est également précisé à quelles échelles ou plages d'échelles les couches de données doivent être affichées (fig. 29).

```
buffer: 16  # Emprise des tuiles élargies d une zone tampon de 16 pixels
srid: 3857  # Définition de la projection. La Web Mercator standard ici
layers:     # Début de l'annonce des couches à intégrer aux tuiles
- name: routes_12_15   # une couche de routes
  queries:
    -minzoom: 12      # Affichée à partir du zoom 12
    -maxzoom: 15      # Jusqu'au zoom 15 compris
    sql:|-
     SELECT *
     FROM
     routes           # Requête sur une table contenant toutes les routes
     WHERE
     geometry && !box!
     AND road_level BETWEEN 12 AND 15 # Un filtre spatial et un filtre attributaire, la tuile contient
                                      uniquement les données de la zone qu elle couvre plus la zone
                                      tampon qui l'entoure et ne contient que les objets de la bases de
                                      données qui portent des valeurs entre 12 et 15 dans le champ
                                      road_level créé au préalable.
```

Figure 29 Exemple de paramètre de création de tuiles avec Utilery contenant une couche de routes pour la plage d'échelle 12 à 15 et s'appuyant sur un prétraitement de la donnée : le champ road_level (Hayat, 2019)

```
tables:                # Début de la liste des tables
 roads :               # Table des routes
  fields:              # Les champs de la table
  - name: osm_id       # Champ portant l'identifiant unique
    type: id
  - name: geometry     # Champ portant la description de la géométrie de l'objet
    type: geometry
  - name: valeur       # Champ portant la clé principale décidé dans le "mapping"
    type: mapping_value
  - key: name          # Champ portant le texte (string) enregistré dans la valeur de la clé name si
                         la valeur existe
    name: nom
    type: string
  - key: bridge
    name: pont
    type: boolint      # La valeur associée à la clé bridge est censée être contribuée sous une forme
                        booléenne (0, 1 ou 'yes', 'no'), Imposm3 interprète ces valeurs
  - name: tags         # Champ qui contient toutes les clés décrivant l'objet
    type: hstore_tags
  filters:
   exclude_tags:     # On ne souhaite pas de routes (highway) qui decrivent une zone (comme des places)
   - -area
     -'yes'
  mapping:
   highway:
   -_any_              # On retient toutes les valeurs possibles associées à la clé highway
   type : linestring
```

Figure 30 Exemple de paramètres de chargement pour une table de routes dans un SGBD PostgreSQL/GIS via Imposm3 (Hayat, 2019)

Le second outil, Tilelive, est un module basé sur Node.js. qui permet de créer un jeu de tuiles vectorielles et de les réunir en une archive MBTiles. Cet outil est un moyen de produire des tuiles vectorielles, en respectant la structure des données visée, tout en s'affranchissant de la question des performances d'interaction entre le client et le

serveur. La structure de données de la base de données source PostgreSQL/PostGIS (chargée avec Imposm3, fig. 30) peut ainsi être différente de la structure de données des tuiles vectorielles (créées avec Utilery ou Tilelive). La base de données source peut donc être utilisée pour diverses utilisations, comme pour la cartographie Web vectorielle et la cartographie papier.

3.1.1.4. De la charte graphique au rendu en ligne

L'organisation des tables de données dans les tuiles fait intégralement partie de la méthode de création cartographique. Les choix de sélection et de hiérarchisation relatifs à la thématique de la carte sont en partie définis dans les paramètres de création des tuiles. Chaque table de données, de la base source, peut être filtrée ou l'ensemble peut être réorganisé en une autre structure, adaptée aux besoins de la charte graphique et de la représentation visée. Dans le cadre de l'expérience d'utilisation de cette technologie, l'équipe géomatique Michelin a créé deux cartes thématiques axées sur la mobilité des piétons en ville et sur la randonnée : deux chartes graphiques ont donc été conçues pour les mêmes tuiles vectorielles.

La méthode et l'approche de cette cartographie seront, en partie, modifiées par rapport à celles utilisées pour la production de cartes en tuiles raster. La structure de la base de données a un rôle clé dans le respect de la ligne éditoriale : les performances d'affichage sont à mettre en place en fonction de la charte graphique. La définition des chartes graphiques partage plusieurs points communs avec des méthodes de cartographie Web utilisant des tuiles raster. Il a ainsi été possible de réutiliser les règles cartographiques (fig. 31) de la carte raster du site ViaMichelin.fr. Par exemple, les objets cartographiques tels que les réseaux routiers, les points d'intérêt ou les toponymes sont distingués et hiérarchisés suivant des variations d'épaisseur de trait, la taille du pictogramme ou la casse et la graisse du texte. En outre, les objets sont sélectionnés et classés selon l'échelle de la carte. Dans le cas décrit ici, le changement technique s'est accompagné d'un changement méthodologique qui doit accompagner le changement de source de données, de TomTom à OSM. Nous présentons par la figure 32 la transposition, que nous avons réalisée, de la charte graphique de la cartographie du site ViaMichelin[182], appliquée aux données TomTom de l'origine vers les données OSM. Il

[182] Les couleurs qui symbolisent les intitulés « Autoroutes », « Escaliers », etc. sont reprises de la charte cartographique de ViaMichelin. Les Z sont les niveaux de zooms de la carte Web auxquels les types d'objets sont affichés.

apparaît que le jeu de données routier, commercialisé par TomTom, spécialistes des données routières, est à la fois détaillé et hiérarchisé. En d'autres termes, le modèle des données TomTom prévoit davantage de catégories et de sous-catégories de routes que le modèle OSM. Celui-ci ne prévoit que rarement des sous-catégories, par l'usage de clés secondaires venant compléter la clé `highway`. Cette étape de confrontation des deux jeux de données rappelle les écarts remarqués lors de certaines évaluations de qualité des données OSM par rapport aux données TomTom. Les données routières OSM sont plus nombreuses et plus denses mais se rapportent davantage à la thématique pédestre et surtout au cyclisme qu'à la thématique du routier.

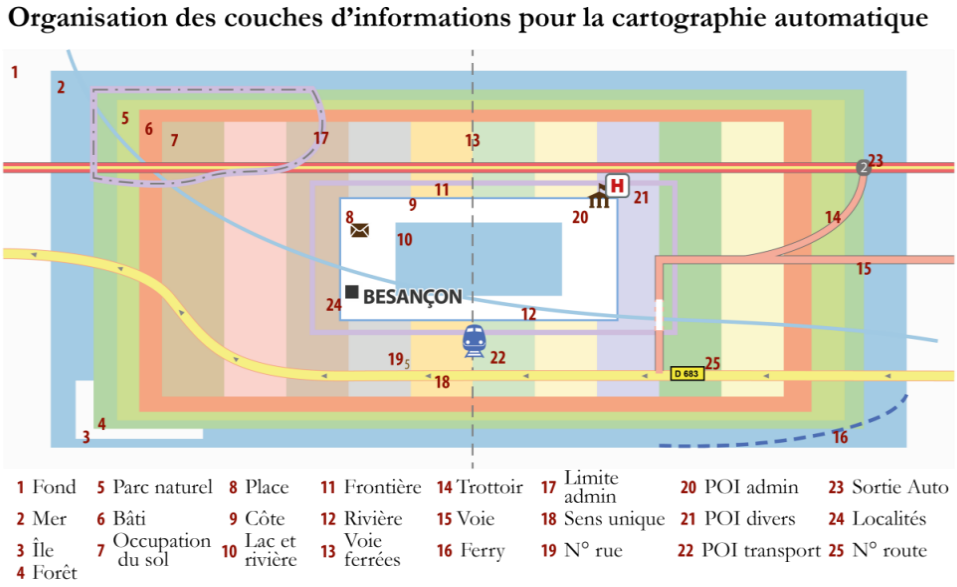

Figure 31 Ordre d'affichage des couches de la carte ViaMichelin (Hayat, 2019)

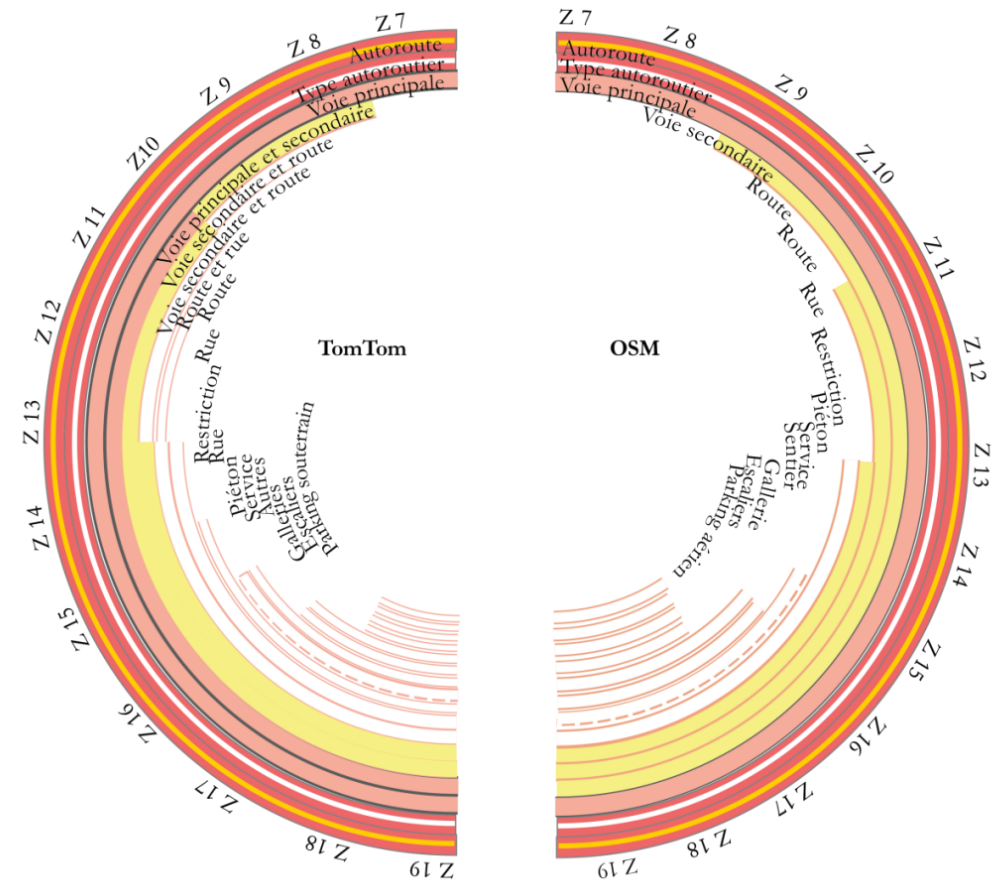

Figure 32 Première conséquence graphique du changement de source, TomTom et OSM ne présentent pas la même finesse de hiérarchisation des routes

Ces règles de cartographie peuvent être appliquées de manière similaire pour la réalisation d'une carte en tuiles vectorielles. Néanmoins, pour produire une telle carte, la charte graphique doit être définie selon la nouvelle utilisation des niveaux de zoom. Il est possible de définir des représentations qui progressent quand l'utilisateur modifie le zoom. Par exemple, l'intensité de la couleur de bâtiments spécifiques peut être progressivement renforcée. Cela permet d'accentuer son rôle structurant dans l'image, à des échelles locales, dès que l'utilisateur agrandit l'échelle de la carte, faisant apparaître ainsi davantage de détails. Enfin, il est important de noter qu'une partie de la sélection des objets est réalisée par la bibliothèque Mapbox GL JS, qui adapte la quantité d'informations à afficher, en fonction de la taille de l'écran et selon des règles que le cartographe définit (fig. 33).

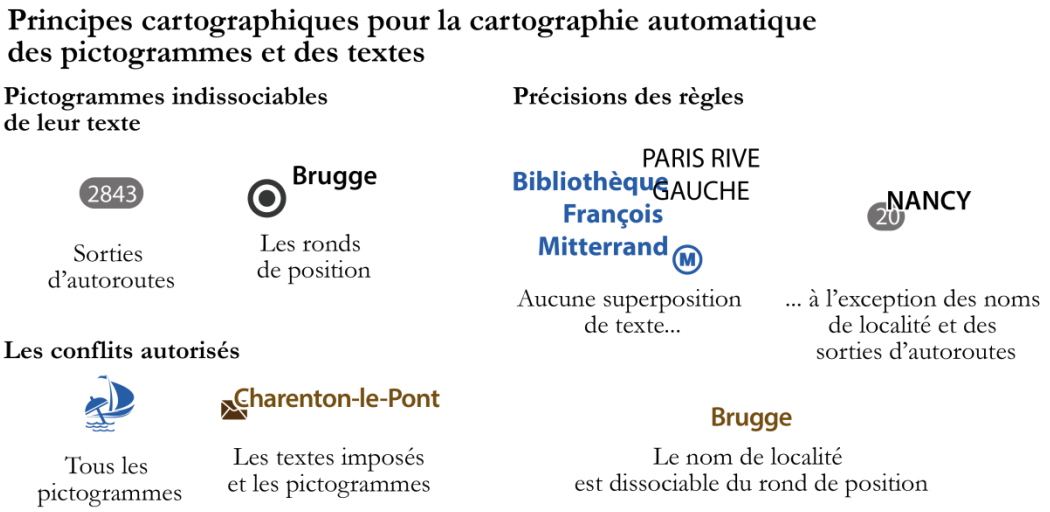

Figure 33 Règles définies pour la cartographie automatique

Les chartes graphiques (fig. 34) peuvent être rédigées comme des paramètres graphiques dans un document de style (JSON). Dans l'exemple présenté, ce document est utilisé par la bibliothèque Mapbox GL JS pour dessiner les données suivant l'ordre, les règles graphiques et les règles d'échelles définies dans ce document.

```
{
    "version": 1,
    "name": "exemple",
    "sources": {                                    // déclaration de l'emplacement des sources des
                                                       données à représenter
        "exemple": {
            "type": "vector",
            "tiles": [ "/all/{z}/{x}/{y}.pbf" ]     // Dans ce cas, une seule source, des tuiles
                                                       vectorielles
        }
    },
    "sprite": "/sprites/sprite",                    // Déclaration de l'emplacement des pictogrammes
                                                       (sprites) et des polices de caractères (glyphs)
    "glyphs": "/glyphhs/{fontstack}/{range}.pbf",
    "layers": [                                     // Début de la représentation des données en couches
        {
            "id": "fond",                           // première couche de représentation, le fond
            "type": "background",
            "paint": {
                "background-color": "#ddff11"
            }
        },
        {
            "id": "espace_en_eau",                 // Deuxième couche, les espaces en eau
            "type": "fill",                        // Des polygones
            "source": "exemple",
            "source-layer": "espace_en_eau",       // Dont la source est la couche de données "espace en
                                                       eau dans les tuiles vectorielles
            "paint": {
                "fill-color": "#1170ff"
            },
          "minzoom":7                              // affichés à partir du zoom 7
}
```

Figure 34 Extrait d'une charte graphique retranscrite en feuille de style

À ce stade du prototype, il ressort que l'aspect le plus intéressant de la technologie vectorielle est le stockage des données dans le cache du client. D'une part, pour un même ensemble de données (dans une tuile) et une plage d'échelles, le client peut naviguer dans une tuile à différentes échelles, sans avoir à télécharger de nouvelles données. L'utilisateur bénéficie d'une navigation fluide dès que les données sont en « cache » et ne consomme pas davantage de données mobiles. D'autre part, l'utilisateur dispose de données sur son appareil et non seulement des images fixes. De ce fait une interactivité est possible hors de toute connexion à un serveur via Internet qui renverrait des réponses à des requêtes. Avec le perfectionnement des appareils, l'utilisateur dispose de jeux de données sous la forme de tuiles vectorielles et d'une capacité de calcul et d'interactivité avec celles-ci directement en local. Cette interactivité possible peut alors s'appliquer par du calcul d'itinéraire, de l'annotation par *mashup* ou même de l'édition de la charte graphique sans aucune connexion au serveur de tuiles cartographiques. Pour un même ensemble de données, le client peut choisir de modifier la charte graphique de la carte. Les attentes graphiques ne sont pas encore totalement satisfaites et l'obstacle majeur reste néanmoins la simplification des géométries aux

petites échelles. Les traitements préliminaires sont donc impératifs pour produire une carte Web qui corresponde aux attentes graphiques de Michelin.

3.1.1.5. Productions papier

Le second prototype[183] est un test de faisabilité de production d'une carte touristique de l'Île de La Réunion à l'échelle 1/80 000. Un extrait géographique de la base de données mondiale (*osm.pbf*) est d'abord téléchargé. Les données sont ensuite chargées dans une base de données PostgreSQL/PostGIS, suivant le modèle de données conçu pour le prototype de la carte Web. Les données peuvent alors être traitées. La majeure partie du processus consiste à traduire les données OSM afin qu'elles soient intégrées dans une infrastructure SIG Michelin. Il devient alors possible d'appliquer automatiquement la charte graphique paramétrée en amont.

Cette étape de remodelage diffère largement de celle qui est conçue pour le prototype de la carte Web. On s'autorise notamment un temps bien plus long, consacré aux développements et à l'exécution des traitements de la base de données, pour les cartes papiers que pour les cartes Web. En effet, dans le cas des cartes Web, l'objectif est de mettre très fréquemment à jour les données dans les tuiles vectorielles. Le traitement des données doit donc être rapide. Dans le cas des cartes papiers, traiter les données a pour conséquence de réduire le temps de travail manuel. En outre, le procédé de création de cartes Web permet l'amélioration progressive de l'esthétique et des choix éditoriaux. Les contraintes de publication sont largement réduites par rapport au support papier, qui demande la réimpression d'une nouvelle édition de la carte et éventuellement le retrait des cartes en vente si le stock n'est pas écoulé.

La production de la carte commence par le dessin automatique de chaque élément dans un logiciel de SIG. Une partie de ce dessin est révisé manuellement. Par exemple, comme la base de données OSM est conçue à très grande échelle, le réseau routier d'une carte à moyenne échelle doit être simplifié et sélectionné pour produire une image claire. On commence par fusionner les routes automatiquement. Les routes à deux voies et les autoroutes sont principalement simplifiées en une seule ligne (fig. 35).

[183] Mis en place dans le cadre de mon travail au sein de l'entreprise Michelin, en 2017.

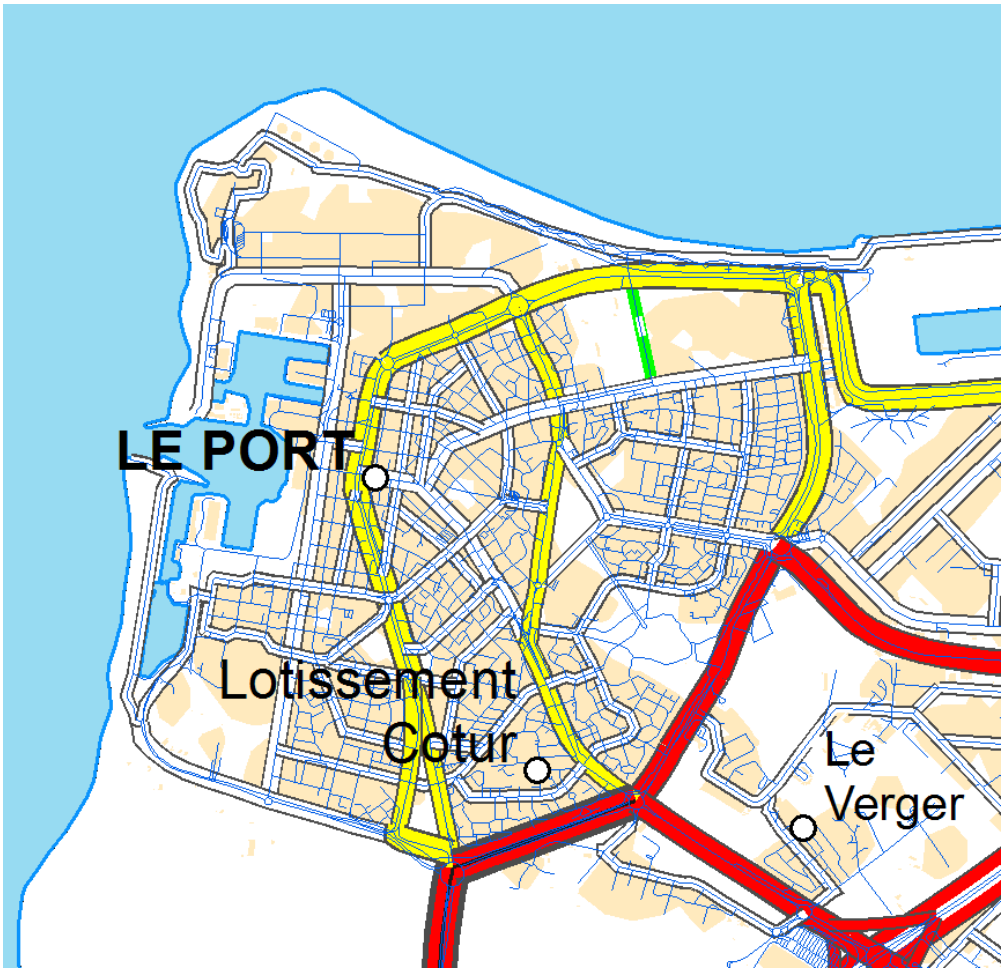

Figure 35 Sélection des routes (en bleu, les tracés initiaux, en rouge les voies principales, en jaune les voies secondaires et en blanc les voies mineure selon une mise en correspondance automatique des tags d'OSM et de la charte graphique des cartes Michelin)

Malgré les traitements géométriques, la complexité d'un réseau routier dense requiert la vérification et la révision manuelle par un cartographe. En effet, la représentation du réseau routier par Michelin vise à la représentation d'un réseau hiérarchisé. Cette hiérarchie visuelle répond à deux objectifs, une lecture simple et rapide des itinéraires possibles en voiture et la compréhension de l'organisation du territoire. Au contraire, la hiérarchisation du réseau dans OSM répond à des critères très diversifiés et la catégorisation est soumise à l'interprétation du contributeur. Ainsi, la mise correspondance automatique entre les *tags* qualifiant les tronçons routiers et la charte graphique Michelin n'est qu'une première étape d'une cartographie lisible et utile à la mobilité (fig. 35 et fig. 36). Le réseau routier est donc partiellement réorganisé pour mettre en évidence sa hiérarchie et certains tronçons sont décalés de leur axe central,

ce qui, à cette échelle de visualisation papier, ne se percevra pas mais permet de gagner en visibilité.

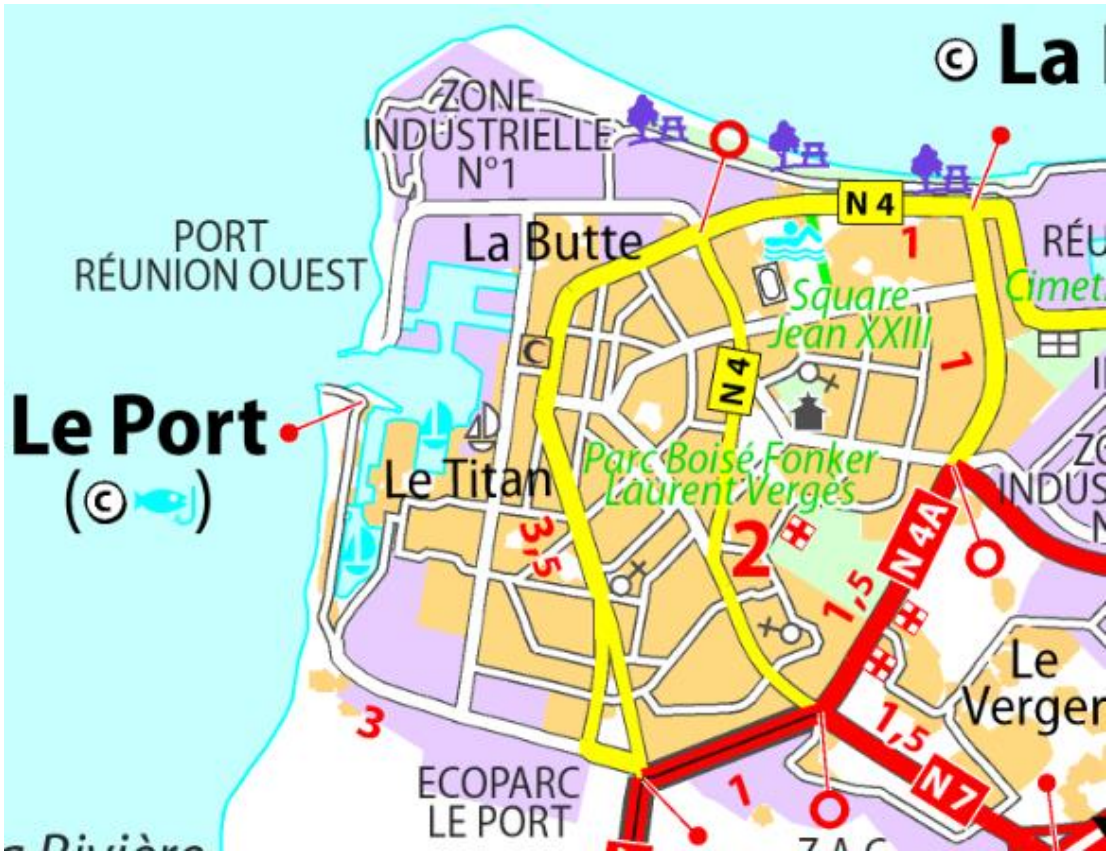

Figure 36 Extrait de la carte finale

Afin de produire la carte de l'île de la Réunion, en utilisant OSM, une enquête de terrain a été également menée sur l'île. Nous menons cette enquête afin de profiter pleinement de l'opportunité de produire des données originales tout en bénéficiant d'un ensemble de données contextuelles déjà créées précédemment par les contributeurs qui nous précèdent. L'enquête est préparée à partir de plusieurs analyses de la base OSM. Une carte de terrain est produite en utilisant principalement les outils automatiques déjà développés pour les prototypes. Sa réalisation complète a révélé des hétérogénéités spatiales et thématiques importantes qui ne peuvent être laissées de côté. Bénéficiant d'un travail de documentation relatif à l'Île de la Réunion, les cartographes enquêteurs ont collecté les informations documentées mais non présents dans la base de données OSM et, en chemin, ont relevé des informations complémentaires inattendues.

Figure 37 Enregistrement d'une trace GPS (tracé bleu), capture d'écran du logiciel JOSM

La collecte de données a été effectuée en utilisant trois moyens de collecte : le GPS (fig. 37), la photographie (fig. 38) et l'enregistrement vocal. Les milliers de photographies horodatées, d'enregistrements vocaux et de notes manuscrites ont été des sources d'information précieuses pour la contribution à la base de données OSM.

Figure 38 Géolocalisation de photos (représentées par un figuré ponctuel symbolisant un appareil photographique) grâce aux relevés GPS (capture d'écran du logiciel JOSM)

Grâce à l'information recueillie et à la pratique du terrain des cartographes, deux thèmes supplémentaires ont pu être développés dans la charte graphique. Tout d'abord, les

routes pittoresques – information propre à la ligne éditoriale de Michelin – ont été identifiées. Ensuite, il est apparu clairement au cours de l'enquête que le pique-nique est, dans la région, une pratique sociale très répandue. Ce thème a donc été ajouté à la carte.

L'activité de recherche chez Michelin a permis d'analyser les différences et les similitudes entre la production d'une carte Web et la production d'une carte papier à partir de données hétérogènes. Les prototypes produits confirment l'hypothèse de la faisabilité technique et graphique, essentiellement pour une carte papier, selon les attentes éditoriales et graphiques de Michelin. Des cartes touristiques Michelin à échelle régionale et des plans de quartiers pour les guides peuvent être créés en utilisant la base de données OSM, grâce à un écosystème technique divers et à une communauté sensibilisée à la question de la réutilisation des données contribuées volontairement. L'ouverture du projet, c'est-à-dire l'accessibilité des données mais surtout la possibilité de les éditer, participe également à la faisabilité de la carte. En effet, la phase d'enquête de terrain qui a précédé la contribution des éléments manquants a été indispensable pour la réalisation de la carte de La Réunion. Ainsi, il a été possible de suivre une convention tacite entre le lecteur de la carte et son producteur suivant laquelle une sélection éditoriale n'est pas dictée par une déficience d'informations. La réalisation des prototypes a consisté en l'intégration d'une pratique nouvelle de la collecte de l'information et de sa mise en carte. Les nouveautés techniques et méthodologiques représentent un défi lorsqu'elles doivent être intégrées dans une chaîne de production plus ancienne et adaptées à un système de production linéaire désormais traditionnel.

3.1.2 Entre nouveaux systèmes de production et assimilation des systèmes traditionnels

L'intégration des informations libres dans une chaîne de production déjà existante présente des avantages qui ont motivé plusieurs fournisseurs et utilisateurs commerciaux ou d'intérêt public à se tourner vers de telles sources. Dans le cas que nous étudions dans cette section, le choix de l'information libre est issu du besoin d'alternatives aux sources utilisées. Deux grands fournisseurs de données ou de services cartographiques nous intéressent, ce sont TomTom, fournisseur de données pour

Michelin et Google Maps[184]. Au sein de Michelin, notre activité de recherche et de développement de prototypes a d'abord été motivée par la nécessité de disposer d'informations géographiques sur des territoires non couverts par TomTom ou de manière insatisfaisante (comme la Réunion, la Corée du Sud, la Colombie). Par la suite, c'est la liberté éditoriale, très appréciée au temps des premiers prototypes, qui a conduit au renoncement, au moins partiel, à une source de données propriétaire et immuable dans les procédés de création.

La recherche d'alternatives à Google Maps, dans l'esprit des adeptes du Libre, au moins sous une forme potentielle, s'est concrétisée en 2012, lorsque Google Maps a commencé à intégrer des tarifications à ses services jusqu'alors gratuits. Ainsi, par exemple, l'encyclopédie Wikipédia a-t-elle changé le fond de carte de ses cartons de localisation au profit d'OSM. Dans le même temps, les multinationales du domaine ont créé leurs propres plateformes de contribution. Google Maps a conçu Google Map Maker, TomTom a proposé TomTom Map Share et Nokia a diffusé Nokia Map Creator (Neis et Zipf, 2012). Si ces acteurs ont choisi le modèle contributif, ils n'ont pas conservé le principe du Libre. De ce fait, les données produites par leurs utilisateurs et les alertes signalées par ceux-ci, sont demeurées la propriété des entreprises. Pour profiter de leurs productions, les utilisateurs particuliers doivent continuer à utiliser leurs services.

Le choix pour l'entreprise Michelin d'intégrer OpenStreetMap à son système de production, a des impacts techniques, éditoriaux, méthodologiques et de communication. Nous résumons les conséquences d'un tel changement de source d'informations dans la figure 39.

[184] Entreprise depuis 2016 et leader quasi monopolistique des services de localisation et de navigation en ligne

Les conséquences des sources propriétaires ou libres sur la pratique de la cartographie

Conséquences	Écosystème OSM		Écosystème propriétaire (TomTom et Esri)	
	Avantages	Inconvénients	Avantages	Inconvénients
Techniques	• Outils évolutifs • Tendance à l'interopérabilité	• Compétences nécessaire en interne à l'entre--prise	• Service après vente • Fiabilité	• Infrastructures peu souples et peu compatibles avec d'autres
Éditoriales	• Originalité des ressources • ressources libres multiples et en développement	• hétérogénéité spatiale et du modèle	• homgénéité du modèle	• Modèle uniforme ne correspondant pas à tous les territoires • Couverture hétérogène
Méthodolo--giques	• Contribution • Conception d'un modèle de données métier à partir des données OSM • Expérience partagée en ligne sur des blogs ou des forums	• Modèle de don--nées et services fluctuants	• Modèle de don--nées et services validés et déjà intégrés dans des chaînes de produ--ction	• L'ensemble des réalisations s'ap--puient sur des technologies dont les codes sources ne sont pas acces-sibles. Problème de la boîte noire
Communication	• Accès à des com--munautés • Conférences spé--cialisées		• Participation aux conférences internationales	

Figure 39 Peser les avantages et les inconvénients pour une entreprise traditionnelle (Hayat, 2019)

Comme nous le résumons dans la figure 32, renoncer à une source de données propriétaires présente des avantages et des inconvénients. Dans un processus de création où la recommandation éditoriale est la principale valeur de la carte, les données libres sont les plus adaptées. C'est le cas lors de la création de cartes et de plans sur supports papier. Dans le cas de Michelin, la possibilité de se placer en producteur de données est en réalité un retour aux pratiques de cartographies qui ont fait le succès de l'entreprise. Du point de vue d'une production de carte numérique pour une emprise mondiale, l'aspect contributif compte mais il est important de reconnaître qu'un producteur seul, n'est pas en mesure de contribuer des informations relatives à l'ensemble du monde de manière fine et en quantité. Dans le processus de création de cartes numériques, c'est l'actualité des données qui prime et elle peut se faire sur un

délai très court, entre quelques minutes et quelques jours selon les prétraitements que l'on souhaite appliquer aux données sources. Si les données propriétaires peuvent être à jour, les délais de livraison impliquent une période de plusieurs semaines, durant laquelle l'information source est figée dans le temps. Par exemple, au lendemain de l'incendie qui a touché, à Paris, la cathédrale Notre-Dame, les informations selon lesquelles le bâtiment n'était pas accessible, tout comme son parvis (fig. 40), étaient déjà renseignées dans la base de données OSM. Cette information est très probablement intégrée dans les livraisons actualisées des données propriétaires, cependant la perturbation du trafic routier est une information relativement peu durable et pourtant indispensable pour un calculateur d'itinéraire ou une carte routière.

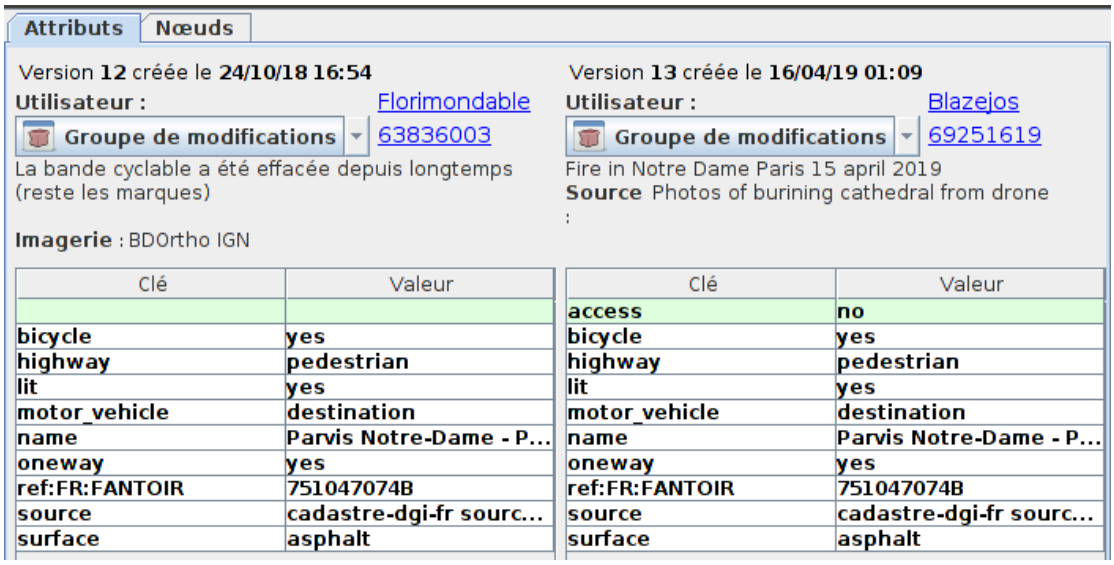

Figure 40 Accès interdit au parvis Notre-Dame renseigné le 16 avril 2019 par l'inscription du tag access=no

Techniquement, le changement de source a des conséquences radicales. D'abord, comme nous l'avons vu dans la précédente section, les outils de chargement des données OSM leur sont spécifiques. Ceux-ci sont publiés sous les conditions de l'*open source*, ce qui représente l'avantage de rendre possible leur publication dans le cas où des modifications auraient été nécessaires. On bénéficie ainsi, à la fois des créations réalisées en interne, pour des besoins métiers, et de la poursuite de l'amélioration des outils, réalisée par les autres contributeurs bénévoles ou acteurs commerciaux. L'appropriation technique de ces outils exige cependant un temps d'apprentissage pour les employés. Ensuite, le modèle de données horizontal et changeant d'OSM est un nouveau défi. De nouveau, l'acquisition de la compréhension complète de sa

construction représente un investissement conséquent afin de l'apprendre, de comprendre son caractère indéterminé et de trouver, parmi une richesse considérable d'informations, celles qui correspondent aux produits à concevoir. Cette étape a représenté plus de deux ans de définition d'un nouveau modèle adapté aux besoins des produits Michelin à partir de la folksonomie d'OSM. Enfin, ce sont des briques techniques de l'infrastructure OSM qui ont dû être réappropriées par les employés de l'entreprise. Si PostgreSQL/GIS et Qgis étaient déjà connus par des employés spécialisés dans le domaine de la géomatique, ce sont des outils ouverts qui ont dû être diffusés plus largement aux cartographes. Par ailleurs, l'utilisation du logiciel JOSM et la pratique du SQL et de l'API Overpass sont trois techniques également nécessaires au bon usage des données OSM. Chacun de ces outils représente des défis d'apprentissage. Le logiciel JOSM et l'API Overpass sont particulièrement complexes à apprendre tant ils sont ancrés dans le schéma OSM.

Pour résumer, choisir une source libre plutôt qu'une source propriétaire pour la production de cartes présente trois intérêts principaux :

- Les cartographes de l'éditeur de cartes sont producteurs. De ce fait, en s'intégrant à une base de données déjà existante, ils sont en mesure d'augmenter simplement les connaissances relatives à un territoire par l'enquête de terrain ou l'utilisation de ressources libres de droit. Ce contexte de création permet l'intégration de spécificités d'intérêt éditorial.

- En bénéficiant de l'activité contributive de la communauté et en y participant, la cartographie numérique publiée est au plus près de l'actualité.

- En s'intégrant à un écosystème ouvert, les employés de l'entreprise mobilisent des outils en évolution et peuvent y contribuer eux-mêmes. L'infrastructure technique mise en place est construite afin de suivre la modernisation des techniques.

Néanmoins, le renoncement à une infrastructure, déjà établie et garantie par des services commerciaux, représente une hausse des besoins en personnel. En effet, il s'agit pour les employés d'apprendre l'usage du système social et technique d'OSM. Il s'agit également d'apprendre à contribuer à la base de données. De nouveau les implications sont à la fois d'ordre communicationnel et technique. En outre, en cas d'erreurs dans les données, de défaillance du système OSM, ne permettant plus, par exemple, le

téléchargement pendant quelques heures, aucun interlocuteur d'autorité ne peut être contacté par l'entreprise, celle-ci n'étant pas liée à OSM par quelque contrat devant être honoré.

Nous présentons sous forme de schémas, deux chaînes de productions différentes par leurs sources de données, qui ont été expérimentées durant notre contrat CIFRE. Les éléments et les étapes qui constituent ces chaînes de production sont simplifiées. Leur représentation vise à expliquer le changement radical de position des éléments de la chaîne quand la source de données est modifiée.

Le premier schéma présente une chaîne linéaire (fig. 41). Les données sont livrées par un fournisseur, ici TomTom, et sont mises en forme automatiquement puis reprises manuellement par des cartographes. La chaîne se sépare en trois produits, les itinéraires, les cartes et plans papiers et la carte numérique ViaMichelin. Si la figuration est simple, l'implémentation a représenté des investissements considérables en termes de temps et de coûts d'infrastructure. Ce n'est pas le détail de cette complexité métier qui est discuté ici mais la position de la source dans la production de la carte. Celle-ci est en début de chaîne et n'est jamais retravaillée. Seule des livraisons de mises à jour réactivent le processus.

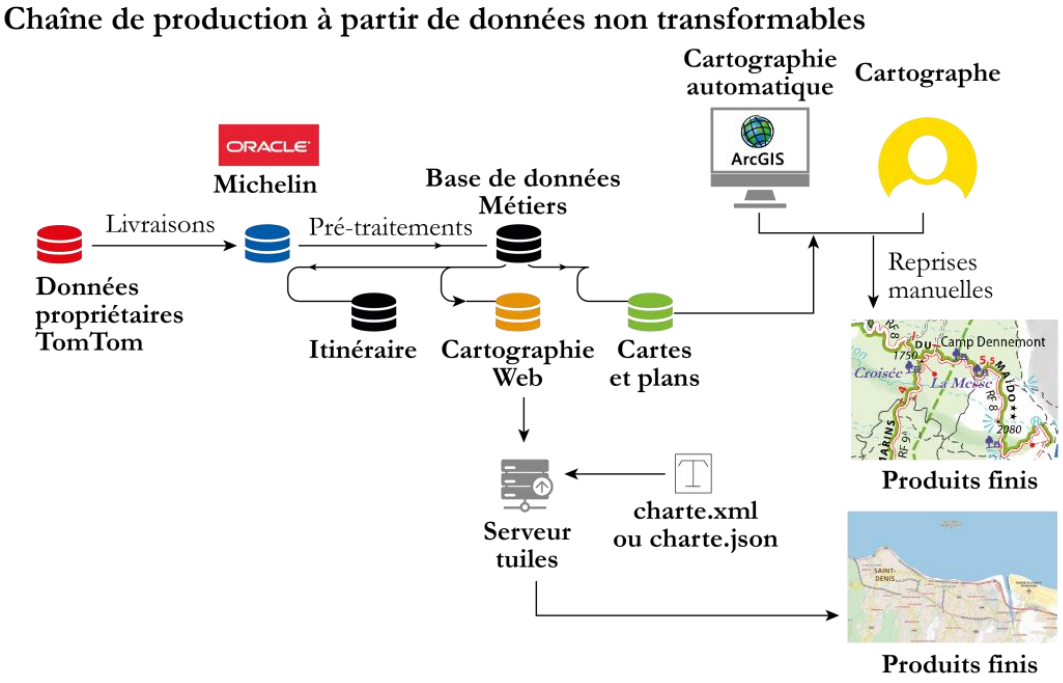

Figure 41La linéarité d'une chaîne de production construite à partir de données propriétaires (Hayat, 2019)

La seconde chaîne (fig. 42) symbolise la synthèse technique réalisée par l'entreprise de cartographie centenaire, dans le but d'intégrer un tout nouveau processus de production d'information. Sur ce second schéma, nous représentons davantage de personnels. L'économie de coût que représente l'utilisation de données libres est transférée sur l'investissement en personnels. Dans le cas d'une création pour support papier notamment, le personnel intervient lors de plusieurs phases de la chaîne, pourtant toujours largement automatique. Après une première création automatique, un prototype de la carte papier est imprimé. Celui sert de base pour la phase de contribution. Les données créées par les cartographes sont accessibles en temps réel et sont, à la fin de la phase de contribution, réinjectées dans la base de données Michelin, qui est une copie sélective de la base de données OSM. À ce stade, les cartographes interviennent de manière plus importante que dans le cas de l'utilisation des données TomTom, pour sélectionner les objets à représenter et parfois même pour généraliser des objets trop détaillés qui auront auparavant été généralisés automatiquement. Par ailleurs, l'activité des cartographes est également nécessaire à la constante redéfinition des processus d'automatisation pour la sélection et la généralisation, tant les territoires représentés dans OSM diffèrent dans leur modélisation et leur finesse de dessin.

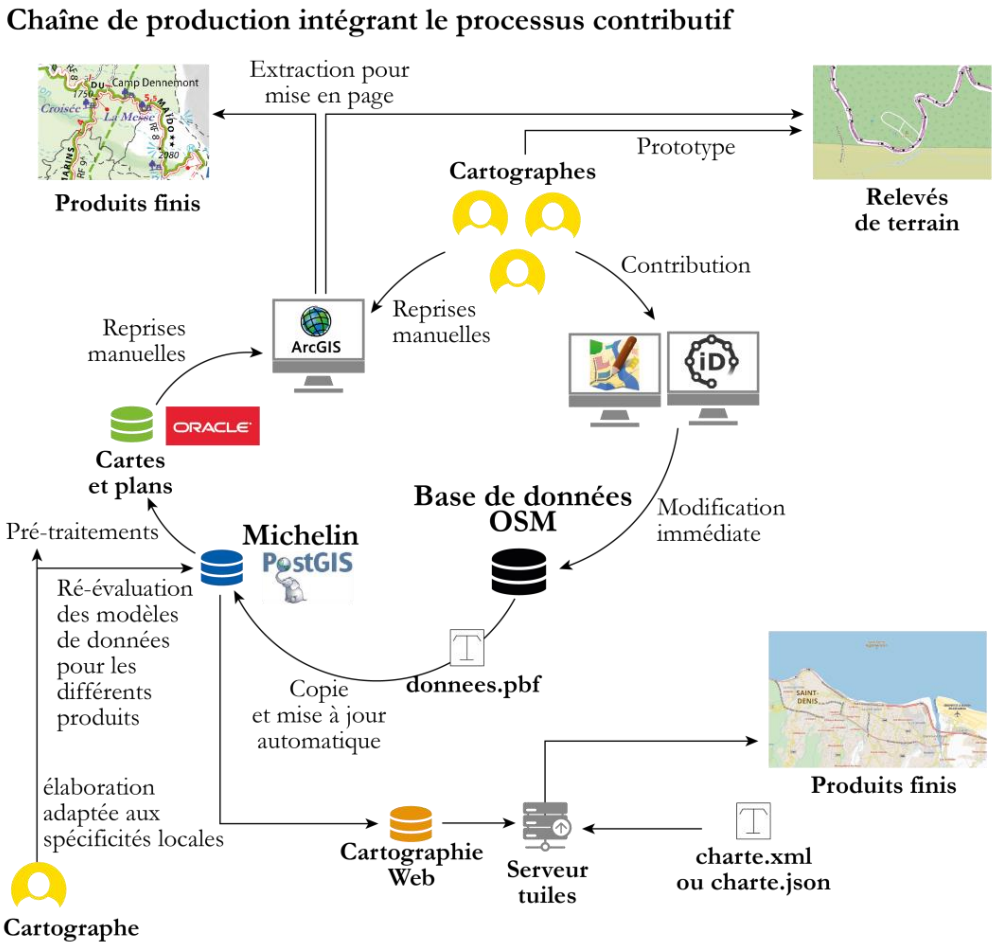

Figure 42 Les changements profonds méthodologiques qui accompagnent l'exploitation complète d'une source libre (hayat, 2019)

Les deux schémas présentent spécifiquement les outils mobilisés, c'est-à-dire le logiciel SIG ArcGIS, les systèmes de gestion de base de données (SGBD) Oracle et PostgreSQL et les éditeurs de données OSM JOSM et iD. S'il est admis que lors d'une réflexion méthodologique, le choix de l'outil relève de ses fonctionnalités générales et non de son fournisseur, il en est autrement dans une chaîne de production industrialisée. Par exemple, les fondements techniques (et commerciaux) du logiciel ArcGIS se mêlent davantage aux fondements techniques du SGBD Oracle qu'à celui de PostgreSQL. Les conséquences d'un changement de SGBD concernent les performances d'affichage, les conditions de stockage et les langages d'écriture de scripts d'automatisation. À la fin de notre contrat CIFRE avec l'entreprise Michelin, une chaîne de production mixte était en place, permettant de progressivement intégrer les outils adaptés à OSM, tout en conservant les choix antérieurs et fonctionnels comme la réalisation complète de cartes et plans par le moyen d'ArcGIS.

Ainsi, nous distinguons différents types d'utilisateurs industriels. Dans un premier temps, la différenciation se reconnaît à l'ancienneté du système sur lequel est fondée la production de cartes. Un acteur traditionnel, antérieur à OpenStreetMap et à la diffusion des SIG *open source*, aura davantage de difficultés à construire un environnement technique adapté à la base de données contributives et libres qu'une entreprise qui a initialement choisi une infrastructure SIG *open source*. En effet, ces nouvelles entreprises intègrent, dès leur fondation, l'écosystème social et technique de projets comme OpenStreetMap en contribuant à son développement. Cependant, aucun fatalisme n'oblige à reproduire cette opposition. Chez Michelin, le système offert par OSM nous a d'abord conduits à reposer régulièrement la question du système de production et à trouver un équilibre entre qualité graphique, gains en dynamisme de publication et mises à jour des données. Dans un second temps, on pourra distinguer les utilisateurs d'informations libres, plus ou moins impliqués dans les activités communautaires. L'implication d'un utilisateur industriel commence par une déclaration, déposée auprès de la communauté productrice, de son utilisation massive des données ou de services mis en place bénévolement. L'utilisateur industriel de l'information géographique se tourne vers une telle source de données dans le but de gagner en autonomie vis-à-vis de ses précédents fournisseurs et de leur modèle de données. Il devient alors, le plus souvent, un acteur de la contribution qui respecte les licences. Les licences jouent ainsi un rôle considérable dans la révélation d'un acteur industriel. En ce qui concerne le domaine du développement logiciel ou algorithmique *open source*, l'utilisateur peut améliorer un outil sans le rendre public, à condition qu'il ne le commercialise pas. Dans ce cas, l'utilisateur ne sera plus en mesure de profiter des futures améliorations de l'outil développé en ligne par une communauté. Il perdra alors l'atout d'un système ouvert. Dans le domaine de l'information et dans notre contexte, les données sont rendues publiques au travers de cartes et de services : il est donc nécessaire, si l'utilisateur souhaite les modifier, qu'il participe publiquement au processus d'amélioration par la contribution. L'utilisateur, qui peut contribuer à de grandes quantités de données, doit chercher à s'intégrer à la communauté.

Comme nous l'avons vu en introduction, la Fondation OSM s'attache à faire perdurer le principe de définition du modèle de données par consensus. Bien que certaines entreprises utilisatrices d'OSM puissent disposer d'une main d'œuvre conséquente et modifier par la contribution quantitative la modélisation de certains objets dans OSM,

le respect du consensus est maintenu. Si la déclaration d'une entreprise utilisatrice et contributrice auprès de la communauté est cohérente avec les principes définis au sein d'OSM, cela pourrait contrarier son activité commerciale. Ainsi, dans un contexte de concurrence et de diminution de la clientèle, dans quelle mesure une entreprise de guides touristiques et de cartes routières et touristiques peut-elle annoncer à ses concurrents sur quel territoire elle contribue ? Une telle annonce anticiperait sur ses futures publications, programmées suite à des analyses marketing. On comprend alors que l'ensemble d'un système de production soit perturbé par l'intégration des principes du Libre. Si les nouveaux acteurs peuvent, au moment de la définition de leur modèle économique, accepter une telle transparence de leurs activités, la mutation d'un acteur déjà ancré dans le domaine exige davantage de réflexion.

3.2. Services et produits cartographiques dérivés de l'information géographique libre

Les services de cartographie en ligne constituent pour l'essentiel le Géoweb c'est-à-dire des moteurs de recherches de ressources sur le Web qui s'effectuent « à la fois par des occurrences textuelles tout en incluant une dimension spatiale (un quartier, une ville, une région, une emprise spatiale, etc.), permettant ainsi la visualisation des résultats de requêtes sous forme de marqueurs sur une carte », (Mericskay, 2016). L'usage de ces marqueurs systématiques dans le cadre du Géoweb fait partie de la pratique dite du *mashup*. C'est à-dire, la création non pas vraiment d'une carte « au sens d'un construit social contenant un message spatialisé sous-jacent » (Mericskay, 2016) mais la représentation spatialisée de ressources enregistrées en ligne, en superposition d'un fond de carte. C'est le choix graphique porté notamment par Google Maps, qui mène à ce que Boris Mericskay (2016) qualifie d'« hégémonie des punaises cartographiques ». Ces punaises ne sont pas des inventions du Géoweb mais un report des utilisations de la carte qui précèdent l'informatique. La punaise d'Internet est une simple analogie de la punaise accrochée à une carte murale. De manière itérative, les utilisateurs d'une carte annotent une carte pour signaler des points d'intérêts. Cependant, l'usage systématique de ces punaises, qui représentent parfois des informations différentes par nature, est caractéristique du Géoweb des années 2010 (Wallace, 2011). Si la tendance que déplorait, T. R. Wallace en 2011, qui consiste à représenter par le même symbole des données pré-traitées et des données ajoutées par des utilisateurs[185], a désormais disparu, en ligne, le figuré ponctuel reste le principal moyen de représenter une information qui n'est pas du domaine du fond de carte mais de la thématique. Esthétiquement et techniquement hégémonique, Google Maps a instauré depuis le début de son service de nouvelles modalités de représentation de l'information géographique.

Aujourd'hui, sur Internet, la vocation principale des cartes interactives est de proposer de très grandes quantités d'informations spatialisées. Les représentations ne sont pas porteuses d'une analyse géographique mais plutôt l'expression de nouvelles compétences de gestion de données disparates qui ont été organisées. Ce nouvel objectif, largement partagé par les diffuseurs de représentations cartographiques, a

[185] Comme le souligne T. R. Wallace (2011), l'usage de la punaise est, depuis la carte murale, le premier moyen d'interagir avec le document cartographique pour l'utilisateur.

mené à une uniformisation du langage cartographique utilisé. En outre, la grande quantité d'informations et la réduction des écrans (issue de la consultation des cartes sur mobile) mènent à des représentations cartographiques pauvres, par figurés ponctuels, représentant chacun une information simple et unique dont la taille varie. L'absence de sélection que cette représentation implique, conduit à une surcharge visuelle de la carte (Loidl et *al.*, 2011). Puisque la carte ne présente pas la synthèse d'une analyse géographique mais la profusion d'informations correspondant à une requête en direct, les « punaises » représentées sur demande ne peuvent symboliser l'ensemble des réponses. La solution choisie n'est pas un retour à la synthèse graphique mais la sélection des informations grâce à des algorithmes. Dans le cadre de services à but lucratif, ces algorithmes sont largement commandés par les revenus que rapporte chaque punaise. Ainsi, si un commerce est un client commercial de Google Maps, sa localisation apparaîtra précocement dans les réponses à une requête lui correspondant. La financiarisation produite suite à la représentation d'un lieu spécifique sur une carte se pratique en dehors de l'entreprise Google Maps. Exploitant les nouveaux moyens d'accès à l'information que représente le Libre, des concurrents de Google Maps développent des services analogues. Cependant, cette pratique de la cartographie en ligne tend à être concurrencée par de nouveaux services commerciaux qui acquièrent désormais des revenus en tentant de proposer des services de cartographies personnalisables. Le langage cartographique en ligne pourrait alors peut-être retrouver une diversité dans la volonté de concurrencer le géant Google Maps.

L'occasion de redistribuer l'usage des services cartographiques, aujourd'hui largement dominés par Google Maps, s'est présentée en juin 2018, lorsque Google Maps a décidé de limiter la gratuité de ses produits. À travers le changement de sa tarification, on comprend quelle était la stratégie de Google Maps. Il s'agissait d'abord de diffuser sur internet un service gratuit et parfaitement fonctionnel. Puis, ce service est devenu progressivement indispensable à de nombreux services commerciaux, qui ont compté sur leur recensement dans Google Maps pour gagner en popularité. D'autres entreprises exploitent la localisation des individus afin de concevoir des services spécifiques, comme Uber.

Différents nouveaux acteurs et acteurs traditionnels du domaine de la cartographie industrielle se tournent vers l'information géographique libre afin de constituer de nouveaux services commerciaux. Comme nous l'avons vu, dans la première section du

chapitre, de nouvelles entreprises comme Mapbox et Maps.me fournissent des moyens de création cartographiques plus diversifiés. Ces entreprises sont également les productrices de produits cartographiques standardisés d'après les normes graphiques de Google Maps. Par ailleurs, de nouvelles plateformes de collectes de photographies des espaces publics reprennent l'innovation que constitue Google Street View. À la différence du service de l'entreprise Google, Mapillary et Telenav mobilisent des contributeurs pour collecter ces photographies. La constitution de ce nouveau système économique est à la fois un accélérateur pour la création d'informations libres (car elle est soutenue par des infrastructures techniques maintenues par des entreprises rentables et stables) et l'émergence d'une nouvelle dynamique perturbatrice pour un système bâti sur le principe du bénévolat (Leon et Quinn, 2018). Ces services s'appuient principalement sur la base communautaire d'OpenStreetMap qui s'est, elle-même, fortement professionnalisée. Nous exposons dans cette section la construction de ces systèmes économiques qui reposent largement sur l'intégration des entreprises aux réseaux sociaux communautaires du Libre.

Cette pratique de l'exploitation d'une main d'œuvre bénévole, qui peut produire de futurs bénéfices professionnels grâce à sa contribution, constitue un nouveau modèle économique dont de nombreux acteurs commerciaux des informations libres sont à dissocier. Des entreprises comme Michelin sont davantage à classer comme des acteurs-utilisateurs et parfois contributeurs au sein de leur entreprise que comme des animateurs des communautés qui dirigent la contribution vers leurs intérêts.

3.2.1. Construction d'un écosystème ouvert et spatial

Outre la technologie GPS dont l'ouverture du service a largement favorisé la production de données spatiales, de nombreux outils, spécifiques au Web ou non, ont permis les utilisations commerciales abordées dans ce chapitre. Il est impossible d'affirmer que, sans ces technologies, aucune organisation communautaire de création et d'utilisation d'informations géographiques n'aurait pu émerger. Cependant, dans l'état actuel de cet écosystème, elles en sont les solides fondations. D'ailleurs, cet état des lieux est provisoire. Depuis 2004, année de la création du projet OpenStreetMap, l'écosystème a évolué ; de nouvelles techniques sont apparues et se sont imposées tandis que d'autres ont été abandonnées. Citons par exemple le passage du SGBD MySQL au SGBD

PostgreSQL, l'abandon de plugins JOSM pour des plugins plus performants, l'abandon également du jeu Kort, une des premières applications visant à la ludification de la contribution à OSM, dont le principe est aujourd'hui repris par StreetComplete.

Ce nouveau contexte de production perdure depuis les années 2000 et repose sur des outils et des bibliothèques qui servent à la gestion de données spatiales (MySQL, PostgreSQL/GIS), à la visualisation de ces informations sous format raster et vectoriel (WebGL-Open GL) et aux interactions client-serveur (JavaScript). La technologie que sert le langage de programmation JavaScript a été favorisée par les capacités matérielles des équipements des utilisateurs particuliers. En 2014, afin de répondre à de nouveaux besoins et dans une simple perspective de transformation et de développement de plus grandes interopérabilité et accessibilité, le langage standardisé HTML a changé de version. Le W3C, valide le HTML5 en tant que nouveau standard. Ce nouveau standard est développé en ayant l'interopérabilité comme principale intention. Cet objectif, commun aux acteurs du développement et de l'adoption du standard, répond à des motivations différentes (Gutiérrez, 2018). D'une part, comme les précédentes versions du standard, le HTML5 répond à l'objectif de construire une structure *open source* pour le Web. Ainsi, la pratique de la définition d'un standard pour un protocole de communication ou un langage informatique suppose la libre disponibilité et l'utilisation universelle du code source. D'autre part, le développement du HTML5 est conduit par des volontés commerciales. Comme l'explique Raúl Tabarés Gutiérrez (2018), plus de dix ans ont été nécessaires à la définition du standard, alors que la tendance du Web de l'époque était à la définition de langages propriétaires non interopérables. Les acteurs les plus concernés par cette problématique de services sur toute plateforme et tout matériel comme Opera, Firefox et Apple ont alors coopéré pour proposer la version HTML5. Raúl Tabarés Gutiérrez (2018) souligne que la coopération s'est imposée du fait du grand nombre d'acteurs du Web. L'hégémonie qui a pu être exercée, pendant une courte période, par Microsoft avec le navigateur Internet Explorer n'est plus envisageable pour aucun des acteurs. Ainsi Raúl Tabarés Gutiérrez (2018) explique que la publication « du HTML5 constitue un tournant majeur dans le développement des standards Web. Les origines de cette innovation reposent sur un groupe de professionnels d'entreprises privées », qui sont parvenus à un consensus dans le but de soutenir l'avenir de leurs services commerciaux.

Le WebGL est une autre technologie associée à l'écosystème permis par HTML5. La technologie WebGL est indispensable pour représenter la donnée envoyée par le serveur, en exploitant le matériel du client. Ainsi le WebGL mobilise la carte graphique de l'ordinateur sur lequel le navigateur ou l'application sont exécutés. En cartographie numérique, le WebGL de même que le HTML5 sont des éléments facilitateurs de la production de cartes. Par sa richesse fonctionnelle, le HTML5 ne nécessite plus de logiciels ou de technologies externes. Cette simplification de la technique facilite la diffusion d'un service cartographique interactif auprès du plus grand nombre. Cependant, afin d'afficher la carte, la transformation continue des technologies du Web exige des utilisateurs qu'ils maintiennent à jour le système de leurs appareils et qu'ils disposent d'un matériel technologique suffisamment performant pour répondre aux besoins de calcul du service affiché dans le navigateur.

Une des limites de diffusion d'un service Web par Internet est la grande diversité des plateformes connectées au réseau Internet, comme les différents navigateurs (Firefox de Mozilla, Chrome de Google, Safari de Apple ou Internet Explorer de Microsoft) et les différents systèmes d'exploitation mobiles (Android de Google, IOS de Apple, Tizen de Linux ou Windows 10 mobile de Microsoft). Pour être consultable, un service publié sur le Web doit être adapté à chaque plateforme. En principe, une des caractéristiques d'un service du Web est d'être interopérable. Les plateformes sont éloignées dans leur conception. C'est pourquoi un service entièrement interopérable est la démonstration d'un investissement considérable pour son développement. Certaines API Web sont développées pour toutes les plateformes du fait de leur succès. Ces services (de données de transport en temps réel, par exemple) sont disponibles, avec des fonctionnalités identiques, quel que soit le type d'applications auquel ils sont intégrés (applications Android, sites Web consultés via Firefox, etc.).

Trois services ouverts de personnalisation de cartes en ligne s'intègrent largement dans le paysage d'Internet.

- En février 2006, la Fondation Open Source Geospatial (OSGeo) a été fondée. En juin 2006, elle a publié OpenLayers, service développé en JavaScript. L'outil réunit une grande palette de fonctionnalités géomatiques permettant, d'une part, la réutilisation de fonds de carte par des sites Web et des transformations géographiques avancées, d'autre part. Cet outil est publié sous licence BSD.

- En 2011, Vladimir Agafonkin de CloudMade[186] (actuellement développeur chez Mapbox[187]) a publié la bibliothèque JavaScript Leaflet sous licence BSD 2-clauses. Simple d'utilisation, l'API s'est diffusée et est utilisée pour l'affichage de la *Slippy Map*[188] du site openstreetmap.org et pour l'affichage de fonds de cartes sur le site de partage de photos Flickr par exemple. Son utilisation est préférée aux deux autres services pour son interopérabilité. En effet, la bibliothèque Leaflet et l'ensemble du service associé ne reposent pas sur les dernières spécifications du HTML 5 mais sur des éléments basiques du langage HTML comme l'affichage d'images (avec la balise <img>). La bibliothèque Leaflet ne recourt pas à la technologie du WebGL. De ce fait, beaucoup de navigateurs, même anciens, correspondent aux attentes techniques du service Leaflet. On peut considérer qu'il s'agit d'une priorité dans la conception de l'outil. En effet, en janvier 2017, dans un *post* du blog de l'entreprise Leaflet annonçant la version 1.0.3 de l'outil, il est fait mention de la version 8 d'Internet Explorer pour laquelle cette nouvelle version était annoncée comme plus performante[189]. La version 8 d'Internet Explorer, publiée en 2009, avait donc déjà huit ans d'ancienneté. Un tel engagement dans le maintien du service Leaflet représente un coût majeur dans le développement de l'outil. Open Layers et Leaflet sont deux moyens d'affichage et de développement de cartes raster interactives. Au contraire de Leaflet, Open Layers est une API qui propose bien plus de fonctionnalités. Leaflet s'est donné pour objectif d'être légère et de fonctionner avec des extensions telles que Géoportail pour Leaflet ou les plugins plus fondamentaux de lecture de CSV ou de dessin de couches vectorielles. Ces fonctionnalités rudimentaires sont intégrées à l'API Open Layers sans avoir à faire appel à des extensions. Cependant, les compétences exigées par l'utilisation

[186] CloudMade est une entreprise fondée en 2007 par Steve Coast et Nick Black. L'entreprise était initialement consacrée au développement d'outils pour le traitement d'informations géographiques et notamment celles issues d'OpenStreetMap. Comme l'indique la page CloudMade du *wiki* OpenStreetMap, après un rachat et le départ de Steve Coast de l'entreprise, CloudMade ne se consacre plus à OpenStreetMap. Depuis 2015, l'entreprise est consacrée aux technologies des voitures connectées.

[187] « *I'm an open source enthusiast at Mapbox, the creator of Leaflet and 40+ other JS libraries, and a rock musician. AMA!* » post de blog sur dev.to en juin 2018 : https://dev.to/mapbox/im-an-open-source-enthusiast-at-mapbox-the-creator-of-leaflet-and-40-other-js-libraries-and-a-rock-musician-ama-1c7o

[188] La *Slippy Map* est le nom donné à la carte présentée par les contributeurs à OSM sur le site openstreetmap.org. Ce nom désigne la carte OSM sur le wiki (Slippy Map, 2018) mais est également le nom type des cartes en ligne composées de tuiles. Ces cartes en ligne sont conçues techniquement pour proposer à l'utilisateur la fonctionnalité de zoom et « faire un panoramique » (*to pan*) sur la carte en cliquant et en glissant à l'aide de la souris. C'est de cette fonctionnalité qu'est issu l'adjectif « *slippy* » ou glissant.

[189] Article Announcing Leaflet 1.0.3, Consulté en ligne en mars 2019 : https://leafletjs.com/2017/01/23/leaflet-1.0.3.html

d'Open Layers dépassent celles d'un utilisateur néophyte. Open Layers exige de l'utilisateur un paramétrage expert comme la définition du système de projection ou du type d'objet géographique qu'il charge (selon les spécifications de l'Open Geospatial Consortium). En outre, l'API n'est compatible qu'avec les dernières versions des navigateurs.

- En 2010, l'entreprise Mapbox est née et s'est positionnée, dès 2011, sur le même type d'offre, en proposant toutefois davantage de fonctionnalités de personnalisation. Jusqu'au développement de son infrastructure de tuiles vectorielles, en 2016 (Mapbox Vector Tile Specification), le service avait comme base la bibliothèque Leaflet. Mapbox s'appuie encore aujourd'hui sur cette technologie pour certains de ses produits de tuiles raster. Ce n'est cependant pas l'essentiel de son commerce, désormais dépendant du WebGL. Pour profiter du service Mapbox, l'utilisateur doit donc disposer d'un matériel récent car le WebGL, à la base du service de tuiles de Mapbox, est exigeant techniquement.

Nous pouvons également citer comme autre API d'affichage de fond de carte et de fonctionnalités interactives, Modest map, développée par Stamen qui est très rudimentaire ou encore, Tangram, de Mapzen. Le service de tuiles vectorielles qui mobilisait la technologie WebGL a fermé en même que l'entreprise Mapzen, en janvier 2018.

Nous présentons là des services qui se déclarent comme appartenant à l'écosystème de l'*open source*. Ce positionnement suppose quelques engagements de transparence et de partage avec les internautes qui peuvent constituer une communauté exerçant ainsi une forme de contrôle moral sur l'entreprise qui bénéficie de l'exploitation commerciale des outils développés en communauté. Néanmoins, l'appellation *open source* ne signifie pas une gouvernance partagée. Lorsque l'écriture de programmes est initiée par une entreprise afin de répondre à ses objectifs commerciaux, une partie des décisions relatives à la fonction, aux langages choisis et à la complexité de l'outil développé sont prises au sein de l'entreprise et sont dénuées de transparence.

La qualification d'o*pen source* pour de tels services repose d'abord sur la reproductibilité du code source, selon des conditions exigeant le moins de contraintes possible. Ainsi, l'exigence de la mention de la paternité, garantit une renommée par la diffusion du produit, tout en ne contraignant pas son utilisation pour d'autres usages

commerciaux. Ensuite, le caractère *open source* nécessite également l'accès au service sans restriction. La déclaration d'une « clé d'API » au serveur de tuiles (rasters ou vecteur), autorisant l'ajout du service de cartographie à un site Web ou à une application, est une restriction forte. C'est ainsi que fonctionnent les services Google Maps ou Bing Maps. En outre, ces services ne publient pas leur modèle de données ni l'ensemble de leur code source.

Cependant, nous considérons qu'une source « ouverte » ne l'est pas uniquement lorsque son code source est lisible et reproductible. Elle est aussi lorsqu'une activité communautaire autour de ce code source est attendue. Parmi ces activités, outre le fait d'intégrer les propositions de contributeurs extérieurs à la production, nous retenons l'imbrication de ces productions avec d'autres projets non affiliés au producteur d'origine. Aux moyens de productions de données, aux logiciels de rendu cartographique en ligne, aux serveurs de tuiles que nous avons déjà décrits s'ajoutent encore d'autres moyens techniques ouverts qui permettent la construction de services. Dans le cas du projet OpenStreetMap, on constate un niveau d'ouverture complet. Ainsi, il est possible depuis le logiciel SIG QGIS d'interroger directement la base de données centrale, via l'API de lecture Overpass. Par l'outil Nominatim[190], la base de données OSM est également mobilisée pour le géocodage ainsi que pour la génération d'adresses potentielles à partir de coordonnées (*reverse geocoding*). Cet usage est possible depuis le site openstreetmap.org ou par le seul usage de Nominatim qui peut être employé avec d'autres sources de données qu'OSM. Pour le même usage, nous retenons l'outil Photon[191]. Plus simple d'utilisation que Nominatim, il permet uniquement de géocoder des lieux.

Un dernier exemple, d'usage potentiel et ouvert de la base de données OSM, est le calcul d'itinéraires. L'outil ouvert Routing Machine (OSRM) permet de calculer les plus courts chemins à partir d'un réseau routier. Il est notamment développé pour être employé en lien avec les données OSM. L'entreprise Mapbox contribue à son développement depuis 2014. Relativement au développement d'OSRM, le

[190] Publié sur le compte Github d'OpenStreetMap, le procédé de contribution proposé est explicité longuement dans une page de site et propose différents modes de contribution entre la déclaration de bugs et la participation aux codes sources : recommandations intitulées « Nominatim contribution guidelines », https://github.com/openstreetmap/Nominatim/blob/master/CONTRIBUTING.md, consulté en avril 2019.

[191] Développé par Komoot, une entreprise de service de calcul d'itinéraire spécialisé sur la pratique du vélo. L'entreprise propose un service gratuit, l'usage public de son API de géocodage et des services payants en fonction de la quantité de requêtes du client.

positionnement de l'entreprise est caractéristique de son utilisation du cadre ouvert, en vue de la production d'outils de haute performance. L'entreprise a ainsi lancé un « challenge » en 2014, via la page Twitter du projet OSRM (fig. 43). Le post renvoie vers un article du blog officiel de Mapbox « Point of interest ». On y lit que « *Mapbox builds—and is built on—open source code. We maintain over 200 projects on GitHub [...]. We are huge supporters of the Open Source Routing Machine and appreciate candidates who understand the benefits of building in the open*. ».[192]

Figure 43 Un défi de développement sur OSRM lancé par Mapbox et diffusé par le compte Twitter d'OSRM

Le calculateur d'itinéraires OSRM est caractéristique de l'écosystème partagé par de nombreux acteurs. En effet, OSRM est un outil central dans la production de services commerciaux basés sur OpenStreetMap. Sur la page Github d'OSRM, on compte trente et un projets auto-déclarés dont des applications en ligne, des librairies de calcul d'itinéraires ou des outils intégrés d'aide à la gestion de flottes pour les livraisons de produits par exemple[193]. Comme souvent, la page Github est incomplète. Si Maps.me n'y est pas mentionnée, elle déclare pourtant utiliser OSRM dans la rubrique « tous droits réservés » de l'application.

[192] Mapbox construit - et est construit sur - du code source ouvert. Nous gérons plus de 200 projets sur GitHub [...]. Nous sommes de grands partisans du calculateur d'itinéraire Open Source (OSRM) et recherchons des candidats qui comprennent les avantages de la construction à ciel ouvert. Traduction par l'auteure. Article Join the Mapbox Directions team publié sur blog.mapbox.com, https://blog.mapbox.com/join-the-mapbox-directions-team-a46dd58ababf, consulté en avril 2019.

[193] Recensement intitulé « Websites and projects using OSRM » sur la page du projet OSRM, https://github.com/Project-OSRM/osrm-backend/wiki/Websites-and-projects-using-OSRM, consulté en mai 2019.

Cette section relative aux services et produits cartographiques dérivés de l'information géographique libre nous a permis d'aborder les motivations et les contextes de productions différents à l'origine d'outils qui composent l'écosystème socio-technique d'OSM. Nous retenons neuf outils constitutifs de l'écosystème d'OSM (fig. 44) en tant que conçus pour l'édition de la base de données (iD, JOSM), pour son chargement et son traitement (Imposm3, osm2pgsql et Osmosis) et pour son exploitation thématique (Nominatim, OSR, Overpass et Photon). Ces outils sont développés en ligne et leurs orientations techniques et fonctionnelles sont débattues sur les plateformes de contribution. Ces discussions ont lieu sur la plateforme Github : dans la figure 44, nous exploitons donc les chiffres fournis sur cette plateforme afin de décrire les rythmes de contribution et les principaux intervenants.

Neufs outils de l'écosystème international OSM, représentatifs de la diversité des acteurs bénévoles et commerciaux

Nom	Fonction	Créé et maintenu par[*1]	Date	Licence
iD	• Éditeur de la base de données OSM • Vise un public novice • Logiciel intégré au site internet openstreetmap.org	• Créé par l'entreprise américaine **Mapbox**, une douzaine de contributeurs et **Richard Fairhust**[*2] • 239 contributeurs en avril 2019 • Principalement maintenu par quatre employés-contributeurs de l'entreprise Mapbox • Codes disponibles sur le github d'OpenStreetMap	2012 - 2013	ISC (licence libre)
Imposm3	• Lecture de tous les formats des fichiers .osm • Écriture des données lues au format PostgreSQL/GIS • Programme utilisé en lignes de commandes	• Créé par l'entreprise allemande Omniscale • 16 contributeurs sur Github • Le principal contributeur est un employé d'Omniscale • Codes disponibles sur le github d'Omniscale	2013	Apache licence 2.0 (licence libre)
JOSM	• Éditeur de la base de données OSM • Vise un public expert • Logiciel à installer en local • À ce logiciel peuvent s'ajouter des plugins déve--lopper par des tierces	• Créé par **Immanuel Scholz**[*2] (allemand, secrétaire de la Fondation entre 2006 et 2007) • 4 contributeurs maintiennent le logiciel et se démarquent par leur **organisation en équipe**. • Codes disponibles sur le github d'OpenStreetMap	2005 - 2006	GNU GPL 2.0 (licence copyleft)
Nominatim	• Recherche d'objets dans la base OSM par leur nom et leur adresse (géocodage) • Génère une adresse à partir d'objets spatialisés (géocodage inversé) • Nécessite l'usage de l'outil osm2pgsql • Programme utilisé en lignes de commandes	• 42 contrbuteurs en avril 2019 dont **Sarah Hauffmann**[*2] • Codes disponibles sur le github d'OpenStreetMap	2010	GNU GPL 2.0 (licence copyleft)
osm2pgsql	• Lecture de tous les formats des fichiers .osm • Écriture des données lues au format PostgreSQL/GIS • Programme utilisé en lignes de commandes	• 71 contrbuteurs en avril 2019 • Maintenu principalement par **Sarah Hauffmann**[*2] (allemande) et **Paul Norman**[*2] (canadien, secrétaire de la Fondation de 2015 à 2018) • Codes disponibles sur le github d'OpenStreetMap	2008	GNU GPL 2.0 (licence copyleft)

Nom	Fonction	Créé et maintenu par[*1]	Date	Licence
Osmosis	• Traitements massifs des données OSM • Chargement d'un fichier XML en une base de données PostgreSQL/GIS	• Créé par **Brett Henderson**[*2] (australien) • 41 contributeurs en avril 2019 • Principalement maintenu par Brett Henderson • Codes disponibles sur le github d'OpenStreetMap	2007	Domaine public
OSRM	• Calculateur de plus courts chemins dans un graphe routier	• 101 contributeurs en avril 2019 • 3 contributeurs de **Mapbox** sont fortement impliqués dans son développement • Codes disponibles sur le github d'OSRM	2013	Apache licence 2.0 (licence libre)
Overpass	• Programme de requêtes de la base de données OSM • Requêtes en lecture	• Créé et maintenu par **Roland Olbricht**[*2] (allemand) • Codes disponibles sur le github de Roland Olbricht	2008	AGPL 3.0 (licence copyleft)
Photon	• Recherche d'objets dans la base OSM par leur nom et leur adresse (géocodage)	• Créé par l'entreprise allemande **Komoot** (service d'itinéraires pour la randonnée pédestre ou à vélo) • 26 contributeurs dont un princi--pal contributeur employé de Komoot • Codes disponibles sur le github de Komoot	2013	Apache licence 2.0 (licence libre)

[*1] Lorsque l'information est disponible en ligne nous indiquons le nombre des contributeurs et nous soulignons les quelques plus importants contributeurs (classés sur Github).

[*2] Nous citons six individus. Ceux-ci sont cités, et leur nationalité est indiquée lorsqu'elle est connue, car nous ne connaissons précisément leur activité professionnelle mais nous avons identifié dans l'importance de leurs contributions la marque de professionnels.

Figure 44 Contributions individuelles ou d'intérêts économiques à l'écosystème OSM (Hayat, 2019)

Nous distinguons trois types d'organisation communautaire :

- **Une organisation communautaire conduite par une entreprise**. Les principaux contributeurs sont des employés de l'entreprise qui administrent les transformations appliquées au code source. L'entreprise conduit le projet, définit les choix techniques et les objectifs fonctionnels. Cette organisation concerne les logiciels iD, Imposm3, OSRM et Photon.

- **Une organisation communautaire resserrée**. Un même groupe restreint ou une personne se charge de l'invention et du maintien d'un outil sur une période relativement longue. Dans cette sélection de neuf outils, trois correspondent à cette description. Ils comptent parmi les outils les plus fondamentaux et les plus originaux de l'infrastructure d'OSM. JOSM, Osmosis et Overpass sont trois outils qui datent du début du projet et qui portent chacun un des trois rôles différents cités précédemment. JOSM permet d'éditer la donnée. À l'aide d'Osmosis, il est possible de charger et de transformer la donnée, tandis qu'Overpass est un outil d'exploration de la donnée, valorisant la richesse thématique de la base de données OSM et ne nécessitant que peu de moyens techniques chez l'utilisateur. Osmosis et Overpass sont deux outils essentiellement garantis par un unique contributeur. Les deux contributeurs que sont Brett Henderson et Rolan Olbricht sont reconnus pour leurs travaux sur la question ; ils contribuent régulièrement et fréquemment à l'évolution des outils. L'organisation autour de l'éditeur JOSM est plus originale. Le logiciel de contribution est celui par lequel le plus grand nombre de données de la base OSM est contribué. JOSM est donc central dans le développement du projet OSM. Une équipe de seulement quatre personnes assure sa mise à jour très fréquente. À la différence de ce qui se passe avec les huit autres outils, l'équipe possède un site Web sur lequel elle présente ses objectifs pour le logiciel. Elle encourage aussi à contribuer les personnes extérieures à l'équipe de même qu'elle incite les utilisateurs à la déclaration d'erreurs.

- **Une organisation communautaire ouverte**. Un programme, conçu par un ou plusieurs développeurs, est mis en commun pour son développement mais également pour définir ses usages possibles. Si quelques contributeurs majeurs se distinguent, la répartition des tâches est plus étendue que dans le cas d'une organisation conduite par une entreprise ou dans le cas où l'organisation est resserrée. Les programmes Nominatim et osm2pgsql présentent cette forme d'organisation.

Enfin, par l'intermédiaire de ces neuf outils, nous avons distingué des personnalités de la contribution logicielle à l'écosystème OSM. D'abord, Sarah Hoffmann, anciennement impliquée dans le comité de l'association OpenStreetMap Suisse, est une des principales contributrices à Nominatim et à osm2pgsql. Elle participe au développement de Photon dans une moindre mesure. Elle partage avec Paul Norman,

secrétaire de la Fondation OSM de 2015 à 2018, le développement principal d'osm2pgsql. Notons qu'Immanuel Scholz, inventeur et développeur de JOSM, s'était porté candidat à l'élection du premier bureau de la Fondation en 2007. Un intérêt pour le bureau de la Fondation a également été manifesté par Rolland Olbricht, inventeur et toujours développeur de l'API Overpass. Pourtant, il n'a pas été élu en 2013, alors qu'il s'était porté candidat. Richard Fairhust, membre du premier bureau de la Fondation en 2007, un des concepteurs du logiciel d'édition iD, se consacre à la diffusion du projet à destination du grand public. Son rôle dans l'invention d'un logiciel d'édition très grand public s'intègre dans ses actions en faveur de la diffusion du projet OSM, notamment à destination d'interlocuteurs aux compétences techniques modestes. Il est le coordonnateur du site switch2osm.org, recueil de textes en faveur d'une solution libre pour l'accès à l'information géographique et de solutions pour la création de cartes Web à partir d'OSM. Enfin, Brett Henderson paraît plus en retrait que les précédents contributeurs que nous avons cités.

Les personnes ci-dessus mentionnées sont des moteurs de la contribution en ligne. Nous insistons sur leur implication associative afin de souligner les relations tissées entre créations techniques et organisations sociales. Les projets comme OSM ou Wikipédia présentent une complexité qui exige la construction de partenariats entre les contributeurs. Dans cette courte recension d'outils, nous signalons également la nationalité des entreprises, des hébergeurs et des contributeurs principaux. L'intention est de montrer le caractère international du projet OSM et l'importance de l'implication des contributeurs allemands, états-uniens, britanniques, canadiens et australiens. Remarquons que quatre de ces nationalités ont en commun l'emploi de l'anglais comme langue principale du pays. Les Allemands se distinguent en ce qu'ils fournissent la communauté la plus importante en nombre de contributeurs à la base de données OSM.

3.2.1.1 Un ensemble de services

La forte présence du système de l'*open source* dans le domaine des données spatiales s'explique partiellement par l'abandon de cette sphère d'activité par les producteurs de logiciels propriétaires. Il a longtemps été impossible de protéger la confidentialité du code source (Gábor Farkas, 2017). En effet, le Web repose sur les interactions entre un client (un navigateur Web mobilisé par un utilisateur) et un serveur (qui délivre des contenus). Les langages de programmation qui servent à développer les logiciels et les

bibliothèques en question ne présentent pas tous les mêmes garanties de protection du code source. Côté serveur, les outils sont écrits en langages compilés. Le code accessible n'est donc pas lisible par un humain. De ce fait, il ne peut être ni analysé, ni reproduit à l'insu de son producteur (reverse engineering). À l'inverse, le JavaScript, langage de programmation adapté aux navigateurs, est plus facilement reproductible, bien que certaines protections aient été développées (Wang and Wu 2014, Gábor Farkas, 2017).

D'autres éléments peuvent venir compléter l'explication du fort développement de l'ouverture des codes sources dans le domaine des données spatiales. On notera d'abord un mouvement général de libération des codes et des jeux de données dans la recherche. Rappelons que l'initiateur du logiciel libre (*free software*) pratiquait son métier de développeur au MIT durant les années 1970. Lorsque la contrainte de nouvelles licences propriétaires est apparu, dans l'univers alors complètement ouvert de son laboratoire, le MIT s'est finalement décidé à le quitter (Bretthauer, 2001). Le MIT est aujourd'hui diffuseur d'une licence libre non *copyleft*, la *MIT license*. L'université de Berkeley est également un acteur important de la recherche dans le processus d'affirmation de la pratique de l'ouverture des logiciels. De 1977 à 1989, des chercheurs de l'université de Berkeley ont travaillé en collaboration avec le laboratoire Bell (concepteur du système d'exploitation UNIX) et ont développé un ensemble de programmes, la *Berkeley Software Distribution* (BSD), qui consistent en une variante du système d'exploitation UNIX. Le projet est partagé auprès d'autres chercheurs du monde entier, auquel l'ensemble logiciel est vendu (Bretthauer, 2001). Finalement, la première version libre de la BSD a été publiée en 1989. Avec cette version libre, une première licence éponyme est apparue. La suite logicielle est libre et non *copyleft*. Les universités américaines impliquées dans le développement informatique ont donc d'abord produit en dehors de toute législation. Cet état naïf a pris fin lorsque l'informatique et la programmation sont devenues des domaines économiquement rentables. Les années comprises entre 1970 et la fin des années 1980 correspondent au temps du débat et de l'expérimentation entre productions ouvertes et productions propriétaires. Le MIT et Berkeley s'intègrent finalement au mouvement de l'ouverture des codes sources et des données afin de répondre à l'exigence fondamentale de la science qu'est la reproductibilité. Petras et *al.* (2015) défendent que l'ensemble des données utilisées dans une recherche publiée doivent être accessibles afin de permettre la vérification

pleine des résultats et des conclusions. En ce sens, les chercheurs encouragent également l'enseignement des SIG (« *Geospatial science education* ») à partir d'outils *open source*. Ils rappellent le risque de la « boîte noire » des logiciels propriétaires dont l'utilisation implique de ne pas pouvoir analyser le code et les algorithmes utilisés pour des travaux de recherche. Leur positionnement pédagogique consiste à proposer un enseignement à la fois pratiqué par les étudiants à partir des logiciels propriétaires et des logiciels libres. Petras et *al.* (2015) considèrent favoriser la diffusion d'une culture de la transparence auprès des étudiants. En outre, leur enseignement a pour objectif de différencier clairement les concepts des logiciels et de permettre une plus grande diversité dans le panel d'outils disponibles lors des pratiques de recherche. En effet, l'ouverture des codes sources permet d'acquérir de l'autonomie face aux produits propriétaires, de transformer ses propres outils pour ses propres besoins et de disposer d'outils compatibles avec d'autres briques techniques ouvertes (Anderson et Moreno-Sanchez, 2003). Plus spécifiquement, l'utilisation de données géographiques est soutenue par la définition des standards FOSS4G permettant la construction de systèmes d'informations dont le stockage, le partage et l'analyse des données sont interopérables (Moreno-Sanchez, 2012).

3.2.1.2. La chaîne de production de la Slippy Map *d'OpenStreetMap*

La chaîne de production de la *Slippy Map* est un exemple d'écosystème ouvert à visée de production et d'exploitation de données spatiales. À la lecture du croquis décrivant les composants de l'ensemble technique d'OpenStreetMap publié sur le *wiki*, une terminologie spécifique au projet OpenStreetMap est employée, en complément du lexique actuel du traitement de l'information géographique. Sur ce schéma (fig. 45), est décrite la chaîne de production du site openstreetmap.org qu'il faut lire comme suit :

1. Le système de gestion de base de données utilisé est PostgreSQL.
2. De cette base sont extraits des fichiers XML dans un format .osm, correspondant donc à un schéma XML spécifique à OpenStreetMap.
3. Ces données extraites peuvent être chargées dans un SGBD disposant d'une extension PostGIS, via osm2pgsql ou Imposm3 par exemple, deux outils spécifiques à la lecture du fichier .osm et à l'écriture de ces données en objets géographiques.

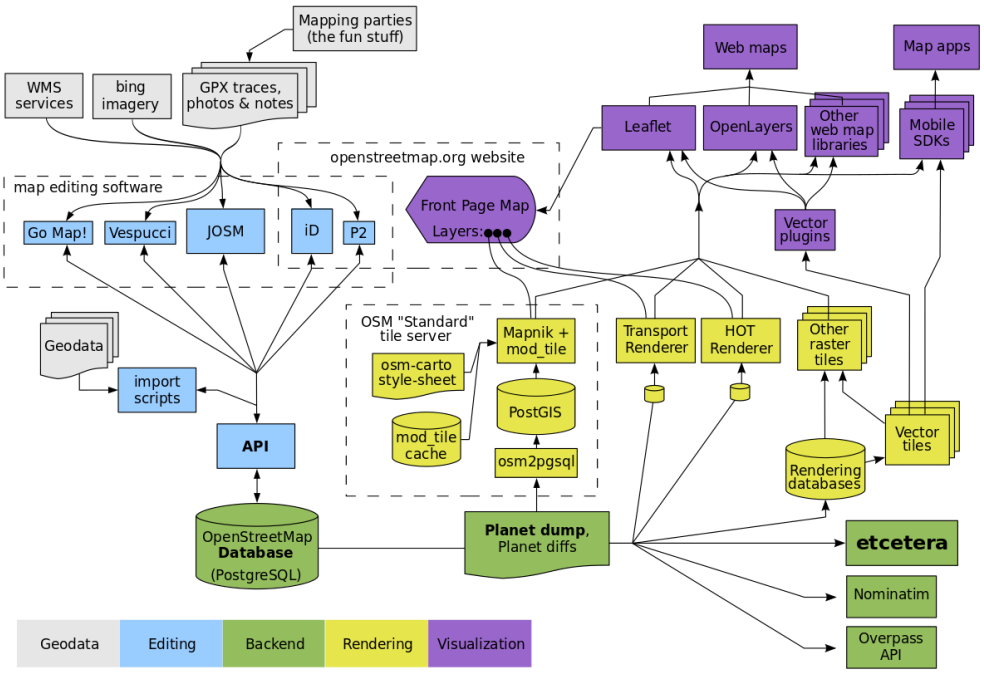

Figure 45 Les composants de l'infrastructure de service OpenStreetMap (Harry Wood, wiki OpenStreetMap, 2017)

À ce stade de la chaîne, il est possible de produire des tuiles raster ou vecteur comme celles de la *Slippy Map* du site openstreetmap.org. Le projet OpenStreetMap s'est appuyé sur le logiciel *open source* Mapnik pour produire ses tuiles. C'est l'API Leaflet qui permet d'afficher les tuiles sur l'appareil du client. Par ailleurs, des outils spécifiques à OSM sont représentés, dans le croquis, reliés au dépôt de la base de données. On trouve Nominatim, outil de géocodage et de *reverse geocoding* (ou géocodage inversé consistant à affecter une adresse à des coordonnées géographiques) et l'API Overpass qui permet d'opérer des sélections dans la base de données centrale du projet en passant par Internet (API). Depuis un navigateur, à l'adresse openstreetmap.org, il possible d'utiliser iD, publié et maintenu par Mapbox, et Potlach 2.

Dans la figure 45, sont représentés également, dans des rectangles verts, les éléments du *back-end* (ou arrière-plan) du projet OSM. Est appelé *back-end*, le socle technique d'une publication, que ce soit un logiciel ou une infrastructure en ligne. Le *front-end* est la partie visible de l'utilisateur. Lorsque la publication est consultée en ligne, le *front-end* est l'ensemble des techniques (manipulées côté client) au travers desquelles l'utilisateur interagit. Comme on le constate, OSM est un projet complet d'une infrastructure pour la création d'informations géographiques en ligne. L'infrastructure

d'ensemble est composée d'un *back-end,* d'un système de production de données géographiques en temps réel et d'un système de publication de cartes en ligne. Le *back-end* s'appuie à la fois sur des technologies et des standards qui perdurent dans le temps (formalismes XML et PBF, enregistrement en base de données PostgreSQL) et à la fois sur des créations originales répondant aux besoins du projet (API Overpass, API d'édition de la base de données, Nominatim). Le système de production d'informations géographiques mobilise les moyens du Web et d'Internet en général. Il s'agit d'un système de production entièrement neuf permettant à des milliers de personnes de contribuer à une unique base de données où qu'elles soient. Cette construction mène Markus Jobst et Georg Gartner (2013) à considérer l'architecture de *back-end* comme le principal facteur du succès rencontré par OSM.

3.2.2. Les producteurs de données, utilisateurs de données, vendeurs de services

Des entreprises privées telles que Telenav, Mapillary, Mapbox ou encore Maps.me produisent de la valeur économique à partir de la main-d'œuvre que représente la communauté des contributeurs. Ces acteurs privés encouragent les contributeurs à participer à un projet à but non-lucratif, sans masquer leurs propres intérêts économiques. Si l'activité contributive bénévole est réelle, elle est régulièrement confrontée à l'ambiguïté de l'utilisation commerciale de ce type de production non rémunérée. L'ambiguïté réside dans l'intérêt concret que représentent ces quatre entreprises pour l'amélioration du projet OSM. En effet, chacune des entreprises, propose soit des plateformes qui soutiennent la production d'informations utiles à la base de données OSM, soit des moyens de contribution directs et simples pour le plus grand public, soit des produits dérivés de la base de données qui prouvent son utilisabilité. C'est dans ce cadre que les contributeurs deviennent une main-d'œuvre bénévole pour des entreprises qui imposent partiellement leurs vues sur le mode de contribution. Dans cette sous partie, nous explorons les causes et les conséquences de l'ambiguïté des pratiques commerciales à partir des données OSM. Les entreprises que sont Telenav, Mapillary, Mapbox et Maps.me sont, selon nous, représentatives de la diversité des acteurs économiques et de leur potentiel international. Telenav est une entreprise traditionnelle du domaine de l'information géographique. Mapillary, présente dans l'écosystème OSM depuis 2014, entretient avec Telenav, tout comme

avec Google Street View, un rapport de concurrence. Mapbox compte parmi les premières entreprises à avoir un système économique basé uniquement sur la concurrence avec Google Maps à partir des données OSM. Si Mapbox est fournisseur de services pour les contributeurs, elle est pourtant décriée du fait de sa visée dirigiste auprès des contributeurs. Enfin Maps.me est une entreprise désormais russe, qui publie une application cartographique hors-ligne pour le grand public. Exemple de la faisabilité d'une production grand public à partir d'OSM, l'entreprise est, quant à elle, critiquée par la communauté pour son service d'édition de la base de données, considéré comme « trop grand public ».

3.2.2.1. Des startups *devenues de grandes entreprises de services*

Parmi les trois acteurs que nous avons choisi d'observer pour leur très forte activité contributive et leurs relations ambiguës avec la communauté. Les trois exemples sont des entreprises nées dans les années 2010, qui ont pris comme socle OpenStreetMap. Ces trois exemples nous apportent un éclairage sur les conséquences industrielles de l'apparition d'OSM en tant que source de données alternative et gratuite. Il s'agit de **Mapbox, Mapillary et Maps.me**.

L'entreprise **Mapbox** a été créée en 2010 par Eric Gundersen. Elle possède deux sièges sociaux, l'un à San Francisco et l'autre à Washington D.C. aux États-Unis. L'entreprise compte plus de deux-cent-cinquante « membres de l'équipe »[194] installés notamment à Washington D.C., San Francisco, Bangalore, Détroit, Berlin, Shanghai, Helsinki. D'une start-up proposant un outil de personnalisation de cartes en ligne publiées sous forme de tuiles raster (flux WMS), Mapbox est devenue, en huit ans, un des leaders dans le développement et la vente de services construits à partir de données géographiques. En 2011, Mapbox diffuse l'outil TileMill, sous licence BSD. L'entreprise s'appuie alors déjà sur d'autres projets ouverts comme Mapnik et node.js. En 2012, Mapbox se détache de l'incubateur *Development Seed*. Elle en fait l'annonce en créant le blog mapbox.com/blog[195].

Au contraire de la Fondation, la société suédoise **Mapillary**, propose des services à la fois gratuits et payants de partage de photographies géolocalisées des espaces publics

[194] Présentation des employés de l'entreprise Mapbox, https://www.mapbox.com/about/team/ , consulté en mai 2018.
[195] Article intitulé « announcing-the-mapbox-blog » sur le blog de development Seed (maintenant reporté sur la plateforme de blogs Medium), https://developmentseed.org/blog/2012/jan/30/mapbox-blog/, consulté en juillet 2018.

du monde, photographies contribuées par des bénévoles. Le service comprend le moyen de contribuer par des photos ponctuelles de lieux remarquables ou des séquences de photographies suivant un parcours dans les rues (figures 46 et 47).

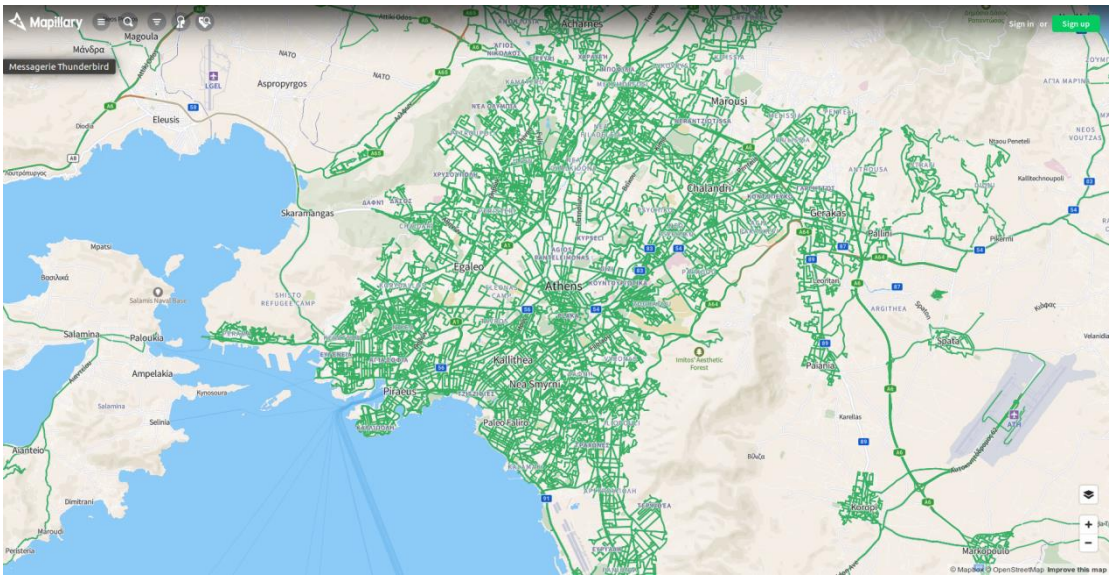

Figure 46 En vert, les tronçons de route photographiés, plate-forme Mapillary utilisant un fond cartographique Mapbox (en vert les tronçons pour lesquels il existe des photographies, en bleu le tronçon sur lequel l'utilisateur s'est positionné, capture d'écran du site https://www.mapillary.com/, en avril 2019)

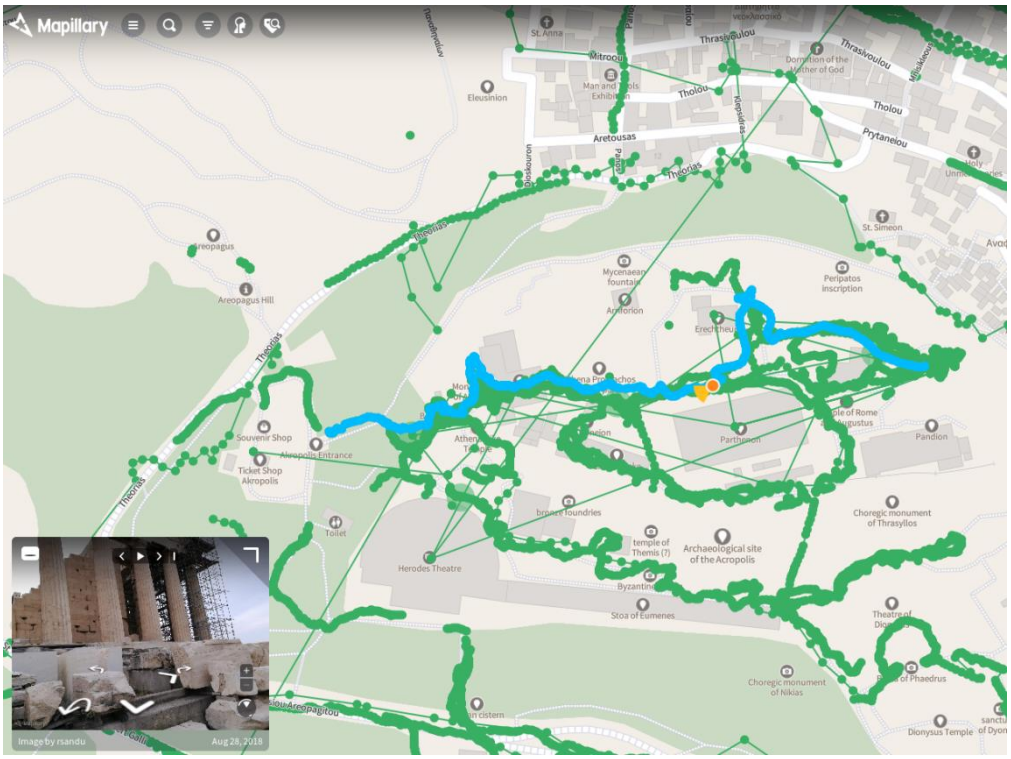

Figure 47 Exploration de la ville d'Athènes via Mapillary (capture d'écran du site https://www.mapillary.com/, en avril 2019)

Mapillary est une entreprise privée ; elle a donc élaboré un service à but lucratif. Les contributeurs photographient les rues via l'application mobile diffusée gratuitement par l'entreprise. L'application peut être utilisée lors d'un déplacement à pieds, en vélo ou en voiture. Lancée en 2014 (avec l'application Android), Mapillary compte en mai 2019 536,1 millions d'images et 7,3 millions de kilomètres parcourus par les contributeurs[196]. La rentabilité du service découle principalement de la vente de services inter-entreprises (B2B) liés aux images (Alvarez Leon, Quinn, 2018). Les services reposent sur des compétences de lecture des images enregistrées par algorithmie afin d'extraire des informations du paysage, des panneaux d'informations routiers, des devantures des magasins, la détection d'objets comme des types de commerces ou du mobilier urbain, etc. La commercialisation de ces images consiste en la vente de services d'expertises, d'abonnements mensuels à la base de données photographiques et de droits exclusifs d'usages des photographies[197]. Les photographies sont, quant à elles, produites par des contributeurs bénévoles. Ce nouveau fonctionnement, issu de l'émergence du contexte du Géoweb, couplé aux notions de l'économie du partage et des communs du numérique, vient perturber un système lui-même récent : celui du contributif libre. Des projets comparables à Mapillary ont le potentiel de changer la dynamique de collaboration qui anime OpenStreetMap et de perturber simultanément l'état de l'écosystème de l'imagerie au niveau des rues, dominé par un seul fournisseur privé : Google Street View (Leon et Quinn, 2018).

Alvarez Leon et Quinn (2018) décrivent également la stratégie économique de Mapillary qui consiste en deux valorisations de la pratique de contribution. D'une part, afin de se distinguer de Google Street View, Mapillary encourage les contributeurs à photographier les zones piétonnes, de voies cyclables et tous les lieux d'accès difficile[198]. D'autre part, l'entreprise articule deux types de propriétés des

[196] Chiffres annoncés sur la page d'accueil du site https://www.mapillary.com/.

[197] Sur son site, Mapillary présente certains de ses clients et le contexte d'exploitation des photographies. Parmi les usages nous retenons, par exemple, la collecte de photographies à but humanitaire par la Croix Rouge, l'exploitation algorithmique pour la reconnaissance des signaux routiers par Mapbox, l'exploitation des photographies pour la planification urbanistique par la ville d'Ottawa ou encore l'intégration des photographies des rues dans la suite ESRI.

[198] Le manifeste de Mapillary confirme ce positionnement : « *Current street view alternatives have so many limitations from the fact that they are created using cars with camera rigs, focusing on streets. There are many places on this planet that they will never get to, and even the places they do get to will not get updated as often as you might want them to. Together we can fix that* ». consulté en ligne en mars 2019 : https://www.mapillary.com/manifesto

photographies. Comme l'indique la page *Licences* du site de Mapillary, en contribuant, le photographe accepte les termes posés par l'entreprise. Les photos sont publiées selon la licence CC-BY-SA. En cas d'usage de la photographie, ouverte à tout utilisateur, les clauses sont donc la mention de la paternité du contributeur et le partage sous des conditions identiques. Le contributeur conserve le droit de supprimer la photographie à tout moment. Par ailleurs, Mapillary possède des droits supérieurs sur les photographies et les métadonnées, selon ce que l'entreprise appelle « une licence commerciale ». Par cette double propriété, Mapillary s'intègre au courant des communs du numérique et construit un nouveau système économique. En outre, l'entreprise s'est assurée d'un partenariat avec OSM en accordant l'accès aux photographies pour les contributeurs à OSM, ainsi qu'à des moyens de détection automatiques des informations relatives au paysage au travers d'outils complémentaires intégrés aux logiciels de contribution JOSM et iD.

Le troisième exemple de service commercial à partir d'OpenStreetMap que nous retenons est celui de la startup **Maps.me**. Son service consiste en une application de cartographie hors-ligne pour téléphones mobiles. Lors de l'utilisation de l'application, il est possible de télécharger des cartes par pays ou par régions. Le système fonctionne à partir de tuiles vectorielles enregistrées sur le téléphone ou la tablette, qui sont affichées à plusieurs échelles sans nécessité de connexion à Internet. L'application offre une interactivité avec les objets des tuiles vectorielles permettant, au clic, d'obtenir des informations complémentaires relatives à l'objet sélectionné, comme le nom de la rue d'un bâtiment ou le numéro de téléphone d'un commerce. Cette interactivité offre également la possibilité de contribuer à la base de données depuis l'application, après s'être authentifié auprès de la base OSM via l'application. La construction de l'entreprise Maps.me est plus complexe à retracer que celles de Mapbox et Mapillary. La page Maps.me du *wiki* d'OSM en décrit une partie (MAPS.ME, 2019). Entre 2011 et 2012, l'application de la startup alors dénommée MapsWithMe commence à être diffusée sur les plateformes de téléchargement d'applications pour mobiles iOS et Android. En 2013, on considère que l'application a été téléchargée plus d'un million de fois. En novembre 2014, le fournisseur d'accès russe Mail.ru annonce son rachat en présentant ce nouveau service comme le moyen de diffusion pour l'entreprise russe de

services de localisation et de mobilité[199]. Sur son site Internet, Maps.me annonce en mars 2019 plus de 120 millions de téléchargements. L'entreprise dispose alors d'un large groupe d'utilisateurs et produit d'abord de la valeur en associant de la publicité aux objets de la carte vectorielle. Par exemple, depuis 2016, Maps.me est associée à Booking.com (service de réservation d'hébergements en ligne) et présente ainsi les services de Booking.com au clic sur un hôtel référencé par celui-ci. L'entreprise diffuse donc gratuitement l'application et propose un service payant de publicité, qui consiste à valoriser les emplacements de ses clients en priorisant le placement des icônes représentant leurs commerces et en les mettant en évidence dans les résultats de recherche d'adresse ou de services par un utilisateur.

3.2.2.2. Des alternatives à Google Maps

Depuis sa création officielle, Mapbox se positionne comme une alternative à Google Maps dans un article de son propre blog « *From Google Maps to Open Source With MapBox and Leaflet* »[200]. Dans cet article, les domaines d'expertise actuellement développés par Mapbox étaient déjà évoqués mais ils étaient alors propres à Google. Mapbox se présentait comme faisant de « belles » cartes, également utilisables comme des fonds de cartes pour des usages commerciaux, aux interfaces rapides et aux tarifs compétitifs par rapport à Google Maps. OpenStreetMap était aussi mentionné. L'accès à cette base de données, couplé aux technologies ouvertes comme Leaflet permettait déjà, selon Mapbox, de proposer des produits cartographiques comparables à la carte de Google Maps. Cependant, dans le même article de blog, il était rappelé que Google proposait, au contraire de Mapbox, des images satellites, des prises de vue dans les rues de nombreuses villes, des calculs d'itinéraires à suivre par des instructions routières et un service de géocodage. On peut ajouter que Google, par le moyen de ses multiples applications, était déjà en mesure de collecter massivement des données géocodées relatives aux activités de ses utilisateurs. Les choix d'investissement technologique pris par l'entreprise Mapbox, née en tant qu'entreprise en 2013, révèlent leur méthode pour concurrencer les services de Google Maps sur l'ensemble des domaines cités dans l'article « *From Google Maps to Open Source With MapBox and Leaflet* ».

---

[199] Communiqué de presse par l'entreprise Mail.ru en novembre 2014. Consulté en ligne en mars 2019 : https://corp.mail.ru/en/press/releases/9209/

[200] Article intitulé « *A Google Maps Alternative: Switch to MapBox* », publié sur le blog de Mapbox, https://blog.mapbox.com/a-google-maps-alternative-switch-to-mapbox-7cc743d3b61a, consulté en janvier 2019.

En décembre 2012, Mapbox publiait une couche de données d'images satellites, MapBox Satellite[201]. Cette couche s'intégrait dans le service de cartographie personnalisée proposé depuis 2010. Elle était fabriquée à partir du catalogue d'images satellites de Landsat et de l'USGS et de l'USDA, toutes sous licences ouvertes. La couverture n'était alors ni mondiale (essentiellement restreinte aux États-Unis), ni exploitable à toutes les échelles. Cependant, cette publication était un premier pas vers une autonomie en ressources documentaires pour digitaliser de nouveaux objets géographiques, déduire des modèles numériques de terrains et ajouter l'esthétique de ces images à son service cartographique. Dans le billet de blog annonçant cette publication, Mapbox assurait qu'une ouverture mondiale complète serait disponible en 2013. Cette offre a été rendue possible grâce à un partenariat avec DigitalGlobe imagery. Depuis son partenariat avec Digital Globe[202], Mapbox dispose d'une grande quantité d'images satellites dont ils sont les ayants droits. Ancien fournisseur de Google en imageries, DigitalGlobe a changé de modèle économique en rendant disponible ses images aux contributeurs OpenStreetMap (DigitalGlobe, 2018), à la manière de Microsoft donnant accès aux imageries Bing. En effet, en novembre 2010, lorsque Steve Coast, fondateur du projet OSM, annonce qu'il va désormais travailler pour Microsoft sur les sujets relatifs à Bing. Il annonce en même temps que Microsoft donne accès pour les contributeurs d'OSM à ses images aériennes[203]. Par ce procédé, Microsoft concurrence Google Maps par le moyen de l'ouverture de ses images satellites et participe à un projet de service cartographique mondial.

En 2015, on trouve la confirmation de l'objectif de Mapbox de concurrencer Google Maps dans un entretien du *Products manager*, Tyler Bell, dans *GeoMarketing from yext*[204]. M. Bell énonce « *We take our brand out of the map so it can be customized for "your" needs. Our message to brands, in-part, is stop letting Google get between you*

[201] Article intitulé « *Introducing MapBox Satellite* », publié sur le blog de Mapbox, https://blog.mapbox.com/introducing-mapbox-satellite-3270de81e45e, consulté en ligne en janvier 2019.
[202] Article intitulé « *Mapping the World with OpenStreetMap via Mapbox* », publié sur le site de Digital Globe, http://blog.digitalglobe.com/news/mappingtheworldinosm/, consulté en ligne en janvier 2019.
[203] Steve Coast écrit sur son blog : «*Microsoft is donating access to it's global orthorectified aerial imagery to help OpenStreetMappers make the map even better than it already is.*», https://web.archive.org/web/20101125151952/http://blog.stevecoast.com/im-working-at-microsoft-and-were-donating-ima, consulté en ligne en mars 2019.
[204] Article intitulé « *Despite Focus In Location Imaging Tools, Mapbox Still Emphasizes The Data* », publié sur le site geomarketing. Com, http://www.geomarketing.com/despite-focus-in-location-imaging-tools-mapbox-still-emphasizes-the-data, consulté en mars 2019.

and your customers »[205]. Trois éléments sont à souligner. Mapbox souhaite se distinguer de Google par son service de personnalisation de cartes. C'est alors le service principal de l'entreprise. Cette personnalisation est destinée à être appropriée par les « marques » à destination des consommateurs. Mapbox signale aux utilisateurs de ses services que Google est un intermédiaire qui profite de ce positionnement. En sous-texte, on comprend que l'information générée par les utilisateurs des cartes diffusées par la plate-forme Mapbox est visée. En effet, Mapbox utilise peut-être déjà, comme Google, sa plate-forme cartographique pour capter les informations géographiques involontaires, qui sont produites à chaque moment de localisation de l'utilisateur utilisant la carte ou à chaque essai de localisation d'objets sur la carte. Tyler Bell précise par la suite tout en confirmant cette analyse : « *So many of our technologies at Mapbox are focused on helping you collect data from disparate sources, making sense of it, updating it quickly, and giving you tools to make it look beautiful. Being able to do that as a developer really opens up a whole new world of interactions of how consumers can engage with your application in exactly the manner you intend* »[206].

La concurrence à Google concerne non seulement son service de cartographie mais également celui des photographies de rue, Street View. Outre Mapillary que nous avons décrite comme proposant un service proche de celui du Street View de Google, l'entreprise Telenav propose un service semblable. Telenav, entreprise fondée en 1999, spécialisée dans les technologies de navigation et de localisation, annonce en septembre 2016 la création de son projet de partages de photographies des espaces publics : OpenStreetView[207]. Le nom OpenStreetView est transformé en OpenStreetCam en novembre 2016. Martijn van Exel, ancien membre du bureau de la Fondation, et porteur de projets relatifs à OSM chez Telenav, publie en novembre 2016 sur le blog Improveosm.org de Telenav une explication de ce changement de nom : «*As part of our fast growing public profile, we have also attracted the attention of Google Inc, who holds the 'Street view' trademark. They are really interested in OpenStreetView but*

[205] Nous retirons notre marque de la carte pour l'adapter à «vos» besoins. Notre message aux entreprises est : cessez de laisser Google se mettre entre vous et vos clients. Traduction de l'auteure.

[206] Un grand nombre de nos technologies chez Mapbox ont pour objectif de vous aider à collecter des données provenant de sources disparates, à les comprendre, à les mettre à jour rapidement et à vous fournir des outils pour un bon rendu. Nos services ouvrent un tout nouveau monde d'interactions sur la manière dont les consommateurs peuvent utiliser votre application exactement comme vous le souhaitez.

[207] Annonce par Telenav sous la forme d'un communiqué de presse. Consulté en mars 2019 : https://www.telenav.com/press-releases/2016-09-22-telenav-releases-openstreetview-an-automotive-integrated-open-source-platform-designed-to-accelerate-the-advancement-of-openstreetmap

also expressed concerns about the name creating confusion. Obviously to us this confusion does not exist, but after considering the pros and cons carefully, we decided to change the name »[208]. Dès le début du projet, la concurrence avec Google Street View est annoncée.

OpenStreetCam est similaire techniquement à la plate-forme Mapillary. Les contributeurs utilisent l'application mobile et photographient les routes et les chemins par son biais. La licence des photographies suit également le même modèle. Les contributeurs sont libres de supprimer les photos qui sont publiées sous la licence CC-BY-SA. mais Telenav dispose de droits supérieurs. L'apport du projet OpenStreetCam pour le projet OSM est similaire à celui de Mapillary en tant que source de données complémentaires pour la contribution. Un *plugin* OpenStreetCam s'ajoute au logiciel de contribution JOSM. Par ce moyen, il est possible de consulter les photographies dont dispose Telenav et certains moyens de reconnaissance automatique sont également rendus accessibles au contributeur. L'usage du recueil de photographies par Telenav ne suit pas la même stratégie que celle pratiquée par Mapbox. En effet, Telenav ne propose pas de services commerciaux qui exploiteraient, à la manière de Mapbox, directement les photographies ou l'expertise qu'elle démontre en matière d'algorithmie de lecture d'images. Telenav exploite les ressources communautaires que sont OpenStreetCam et OpenStreetMap en intégrant les données aux services commerciaux qu'elle publie depuis ses débuts.

Enfin, l'application Maps.me[209] est une application de cartographie hors-ligne, qui se présente comme une alternative aux services cartographiques populaires (figure 48). C'est ainsi que la description qui accompagne l'option de téléchargement de l'application sous Android en langue anglaise précise que Maps.me est : «*UP-TO-DATE - Maps are updated by millions of OpenStreetMap contributors daily. OSM is an open-source alternative to popular map services*»[210]. Nous notons une traduction libre de cette description dans la version française de cette phrase : «À JOUR - Les

[208] Dans le cadre de notre communication publique, nous avons également attiré l'attention de Google Inc., qui détient la marque «Street View». Ils sont vraiment intéressés par OpenStreetView, mais ont également exprimé leurs inquiétudes quant à la confusion avec leur nom. Évidemment, cette confusion n'existe pas pour nous, mais après avoir soigneusement examiné le pour et le contre, nous avons décidé de changer de nom.

[209] Sur le même plan, l'application OsmAnd propose une application également hors-ligne et présentée dans les recherches sur Internet comme une alternative à Google Maps.

[210] Description disponible en ligne ou via l'application mobile GooglePlay, https://play.google.com/store/apps/details?id=com.mapswithme.maps.pro&hl=en

cartes sont mises à jour par des millions de contributeurs à OpenStreetMap sur une base quotidienne. OSM offre une solution cartographique en *open source* en remplacement de Mappy»[211]. Pour le traducteur, c'est l'entreprise de cartographie Mappy qui est concurrencée plutôt que l'ensemble des services prisés de cartographie. Si Maps.me s'attache à ne pas désigner nommément son concurrent principal, ce sont les ressources en ligne proposant des alternatives aux services de Google qui s'en chargent. Par exemple, le site Framalibre « annuaire du Libre », présente l'application cartographique comme une alternative aux applications Google Maps et Here (acteurs des services de navigation embarqués et en ligne, anciennement Navteq et actuellement propriété de BMW).

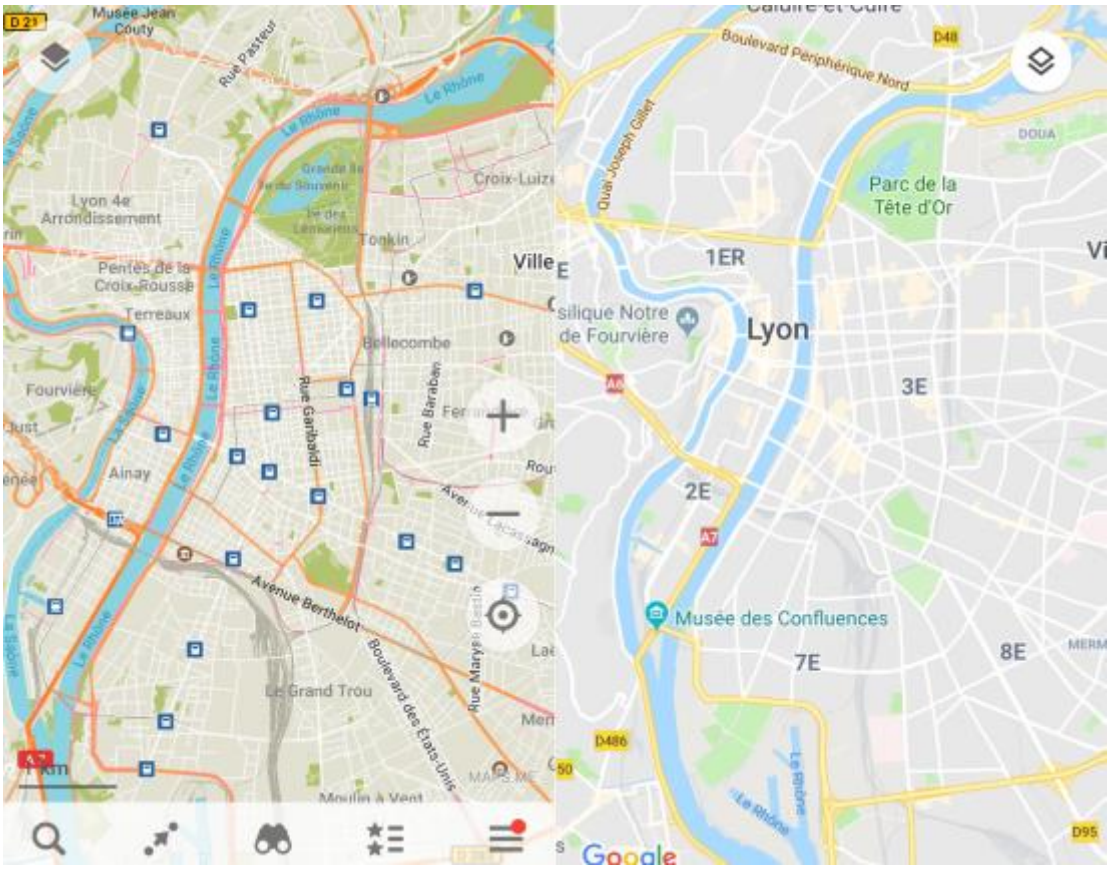

Figure 48 À gauche, la représentation par Maps.me et à droite la représentation par Google Maps

(Capture d'écran des applications mobiles Maps.me et Google Maps, en avril 2019)

[211] https://play.google.com/store/apps/details?id=com.mapswithme.maps.pro&hl=fr

3.2.2.3. Les relations entre les acteurs commerciaux et les communautés

Le système commercial conçu par Mapbox et Mapillary repose sur l'équilibre fragile entre une dépendance forte à la main-d'œuvre bénévole susceptible de se désimpliquer à tout moment et la production de valeur à partir de données contributives. Comme le soulignent Alvarez Leon et Quinn (2018), la valorisation économique de la contribution présente des défis uniques. L'un d'eux est la nécessité de garantir un volume élevé, une qualité constante et une couverture étendue des images soumises par les utilisateurs. Pour résoudre ce problème et encourager les contributeurs, Mapillary utilise un procédé de *ludification* : l'énumération publique des personnes et des organisations qui ont contribué au plus grand nombre de photos du site. Les résultats sont affichés en évidence avec les statistiques des meilleurs contributeurs sur le blog de Mapillary (figure 49). Ce procédé est également choisi par Telenav pour encourager la contribution à OpenStreetCam (figure 50 et 51).

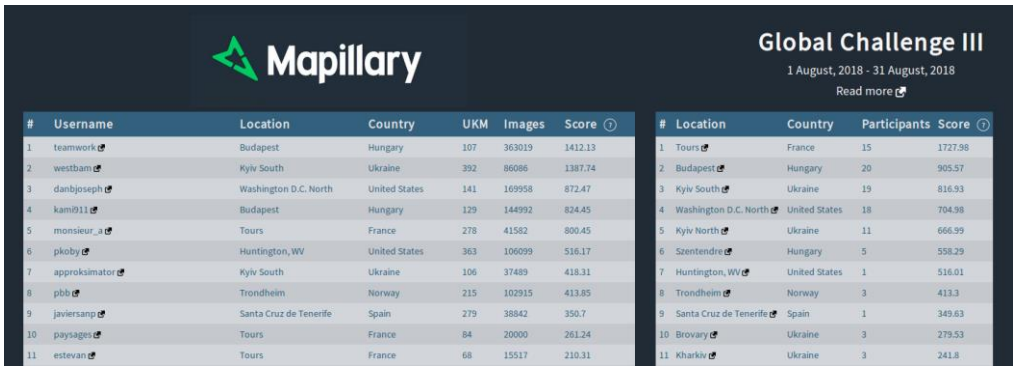

Figure 49 Les meilleurs résultats du challenge #CompleteTheMap par Mapillary durant le mois d'août 2018 (capture d'écran du site https://mapillary.github.io/mapillary_greenhouse/global-challenge/q3-2018/, en avril 2019)

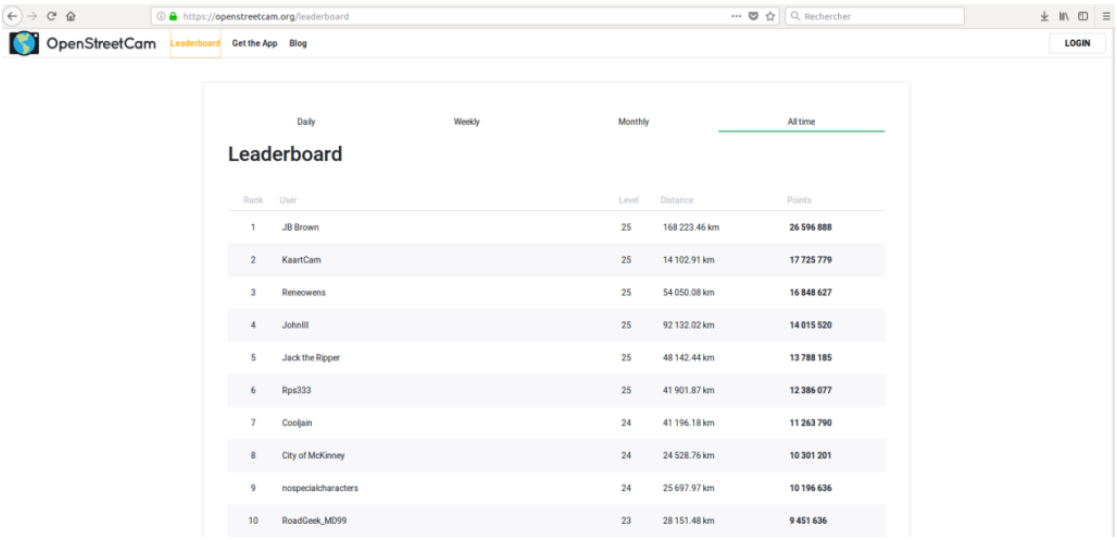

Figure 50 Classement en direct des contributeurs en fonction des kilomètres parcourus et du nombre de photographies chargées sur OpenStreetcam

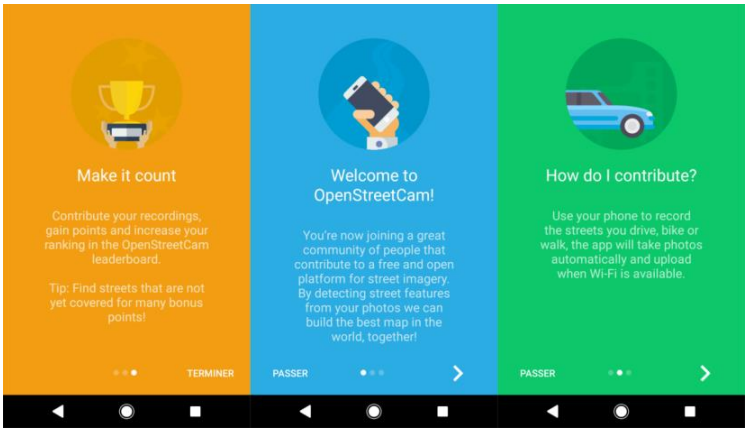

Figure 51 Les trois panneaux de présentation au premier allumage de OpenStreetCam : gagner des points (à gauche), faire partie d'une communauté (au milieu) et la contribution est facilitée par le moyen de l'application (à droite)

Ainsi, OpenStreetCam et Mapillary construisent des communautés de contributeurs similaires qui partagent les mêmes objectifs. C'est d'ailleurs la raison pour laquelle de nombreuses entrées de forum, de posts sur *Twitter* et de pages *Github,* par exemple, discutent d'une éventuelle compatibilité entre les deux projets. Lorsqu'en août 2016 Greg Sterling, auteur d'articles d'actualité relatifs à Internet, publie un article annonçant une alternative à Google Street View, sous le nom d'OpenStreetView, ouverte et associée à OpenStreetMap et le signale sur Twitter, Peter Neubauer, employé à

Mapillary, réagit dès le lendemain et rappelle que son entreprise existe depuis 2014 et possède alors 70 millions de photographies (figure 52).

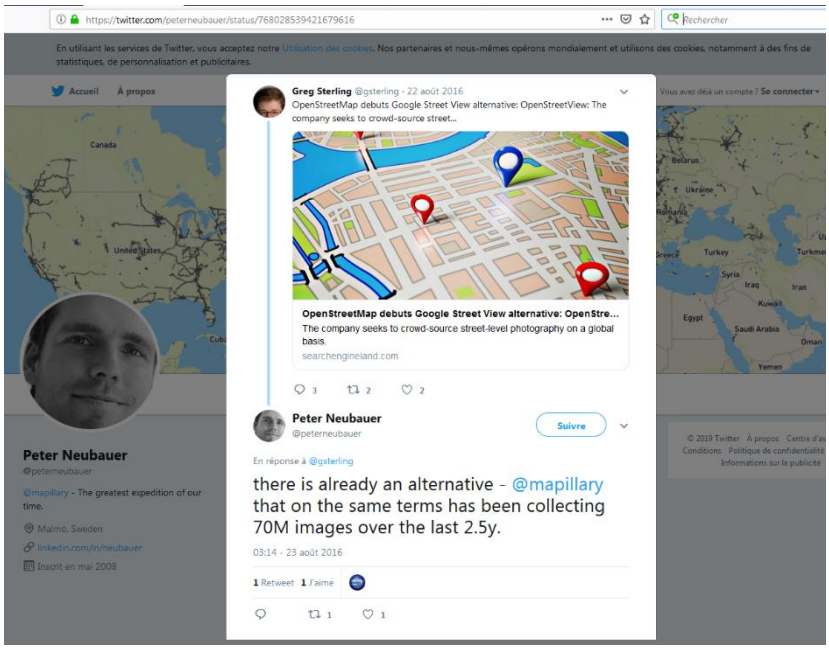

Figure 52 Mapillary manifestant sur Twitter son avance sur OpenStreetView (OpenStreetCam aujourd'hui)

Le 22 août 2016, une discussion (*Issue*) intitulée « Cooperate with Mapillary? / Difference to Mapillary? #60 » est ouverte sur le compte Github d'OpenStreetCam. Le contributeur Rugk ouvre la discussion en signalant que le projet Mapillary propose le même service qu'OpenStreetCam. Il précise : bien que les licences des photographies soient différentes, le projet OpenStreetCam devrait être en mesure d'intégrer, dans sa propre base de données, les photographies enregistrées par Mapillary[212].

Le 22 août 2016, une discussion (Issue) intitulée « *Cooperate with Mapillary? / Difference to Mapillary? #60* »[213] est ouverte sur le compte Github de OpenStreetCam. Le contributeur *Rugk* ouvre la discussion en signalant que le projet Mapillary propose le même service qu'OpenStreetCam. Il précise : bien que les licences des photographies soient différentes, le projet OpenStreetCam devrait être en mesure d'intégrer, dans sa

[212] La discussion intitulée « *Cooperate with Mapillary? / Difference to Mapillary*? #60 » commence par : « *Mapillary basically does the same job as this new project. The images are under different CC licenses, so you should be able to use them and incooperate them into your own project* ». Consulté en ligne en mars 2019 : https://github.com/openstreetcam/openstreetcam.org/issues/60.
[213] Coopération avec Mapillary? / Différence avec Mapillary?

propre base de données, les photographies enregistrées par Mapillary[214]. Le débat s'ouvre à propos d'une possible préférence de contribution entre Mapillary et OpenStreetCam. La duplication d'un espace de contribution de photographies est un problème qui ne se pose pas pour OpenStreetMap, unique dans le domaine des données géographiques. Cette situation est analysée par Leon et Quinn (2018) qui décrivent une relation de « cannibalisation » entre Mapillary et OpenStreetCam, qui se partagent une communauté de contributeurs sans leur permettre une contribution commune aux deux projets. La discussion sur Github oppose donc les deux projets, comparant leurs niveaux d'intégration dans le Libre et leurs objectifs commerciaux. Après une journée de discussion et une dizaine de messages échangés par des individus manifestement contributeurs mais extérieurs aux entreprises en question, un employé de chacune des entreprises se manifeste. Le premier est employé par Telenav : Martijn van Exel. Il est impliqué dans la communication de l'entreprise au travers du blog et de sa page Twitter. Il écrit ici sous son pseudonyme Github (*mvexel*). Il rappelle la différence de visée économique entre les deux projets et soutient l'objectif principal de contribution à OpenStreetMap au travers de la plateforme de photographie. S'il souligne que Mapillary présente alors une infrastructure plus stable et plus élaborée que celle d'OpenStreetCam, il affirme également que Telenav est sur le point de publier des outils encore plus utiles pour le projet OSM que ceux de Mapillary. Le lendemain, un employé de Mapillary répond, sous le nom de *jesolem*, à cette comparaison. Il explique que l'entreprise Mapillary fournit des services techniques aux contributeurs d'OSM tandis que Telenav ne fait qu'utiliser OSM. Enfin, il affirme que Mapillary a proposé un partenariat avec OpenStreetCam (en réponse aux demandes des contributeurs de ne procéder qu'à une contribution unique, qui serait utilisable par les deux projets) qui aurait été refusé par Telenav. Le débat se poursuit sur le *wiki* OSM et reprend les termes de la discussion citée. Martijn van Exel écrit finalement sur la page Talk:OpenStreetCam qu'aucun projet n'est complètement libre ou complétement propriétaire mais qu'OpenStreetCam présente davantage de composants *open source* que Mapillary. Ainsi, ces nouvelles entreprises qui se positionnent comme des alternatives aux produits de Google tentent de se distinguer du géant en se présentant comme ouvertes à la communauté. Certaines de leurs innovations restent pourtant

[214] *Rugk* répond ce 22 août 2016 : *« So I don't really know what is the sense of this project. Mapillary already does it and does it better (from a feature and UX perspective). So why should I, as a user of Mapillary, switch? Or - which would be better - could you just cooperate to integrate the images you collect into both databases*?" Consulté en ligne en avril 2019 : https://github.com/openstreetcam/openstreetcam.org/issues/60

fermées, nous supposons dans l'optique de conserver leurs originalités, mais peuvent alors se placer en contradiction avec leurs déclarations publiques en faveur du tout *open source*.

3.2.2.4. Inciter à la contribution

Si les projets comme Mapillary et OpenStreetCam reposent nécessairement sur des communautés, les acteurs commerciaux utilisant OSM peuvent soit l'exploiter telle quelle, soit inciter à la contribution et parfois même la diriger. Ainsi, l'entreprise Mapbox est motrice dans la production des données par les contributeurs en maintenant l'éditeur de la base de données iD, mis en place par défaut sur le site Openstreetmap.org, dès 2013. En 2014, Mapbox annonce avoir améliorée le logiciel iD[215] sur son blog. Désormais, selon Mapbox, la contribution des conditions de circulation aux intersections (*turn_restriction* dans OSM) est facilitée. Cette information, relativement complexe à ajouter par le moyen de JOSM, est pourtant nécessaire pour les applications de calcul d'itinéraires. Mapbox ne cache pas ses intentions dans ce post de blog et souligne qu'elle propose un service de calcul d'itinéraire qui nécessite la bonne contribution (homogénéisée donc) de cette information. L'entreprise signale également avoir déjà contribué une centaine de *turn_restriction* sur San Francisco. La simplification de la contribution au travers du logiciel iD vise d'une part à toucher le plus grand nombre et d'autre part à uniformiser les contributions. En effet, l'outil est conçu pour que la contribution se réalise en langue courante. Les *tags* sont masqués derrières des appellations comme « route résidentielle », « parc » ou encore « lac ». Ces expressions préenregistrées correspondent à un ou plusieurs *tags*. La contribution est ainsi partiellement guidée par Mapbox qui définit les *tags* nécessaires aux appellations proposées au contributeur novice. Enfin, iD intègre une option permettant d'afficher les photographies enregistrées sur la plateforme Mapillary. Les algorithmes de détection d'informations dans le paysage identifient notamment les informations routières, participant à l'amélioration de la description du réseau routier.

Maps.me est la deuxième entreprise de notre sélection qui construit une boucle vertueuse au profit de ses services, grâce aux moyens de contribution qu'elle intègre à son application mobile. À l'instar de Mapbox, l'objectif de Maps.me est de rendre la

[215] Dans l'article de blog intitulé « Simple Editing for Turn Restrictions in OpenStreetMap », https://blog.mapbox.com/simple-editing-for-turn-restrictions-in-openstreetmap-36c1d6e4bbe9, consulté en ligne en avril 2019

contribution la plus simple possible. Elle vise des contributeurs novices et peu enclins à l'utilisation d'un logiciel spécifique tel que JOSM. L'application est cependant très contestée par une partie de la communauté de contributeurs. Déjà, en 2016, Maps.me s'est vu reprocher son accord commercial avec l'entreprise de réservation d'hébergements en ligne Booking.com. Un fil de discussion a été ouvert sur la liste de discussion Talk OSM afin de discuter de la légalité du procédé choisi par Maps.me. L'entreprise a directement répondu sur le fil afin d'affirmer la légalité de son procédé, dans le respect de la licence ODbL et des préconisations faites par la Fondation. Suite au constat de la faible qualité des contributions généralement réalisées par le moyen de l'application Maps.me, depuis 2017, la communauté fait reproche à l'entreprise de certains de ses choix commerciaux et de contribution. La page du *wiki* OSM relative à l'entreprise présente une rubrique intitulée « Modifications discutables » ou, en anglais, « *Questionable_edits » (*MAPS.ME, 2019*).* À la différence d'iD qui propose un tutoriel à destination des nouveaux utilisateurs, aucune indication n'est donnée au contributeur débutant. En outre, il apparaît que certaines modifications sont appliquées à la base de données sans que l'utilisateur en ait conscience. Certains utilisateurs semblent considérer la partie *contribution* davantage comme un outil pour le *mashup* et la création de cartes personnelles, plutôt que comme un outil de contribution à une base de données publique. En outre, les propositions de *tags* préenregistrés existent mais sont peu nombreuses. La page du *wiki* insiste sur l'usage erroné des *tags* `tourism=guest_house` et `tourism=attraction`, deux *tags* fondamentaux pour le service de tourisme de Maps.me.

Les services commerciaux que nous avons énumérés offrent tout à la fois des solutions simplifiant la contribution et l'utilisation des données de même que des services commerciaux centrés surtout sur la mobilité. Dans le cadre de Michelin, c'est plutôt la ressource documentaire, potentiellement adaptée à chaque territoire, qui a fait d'OSM une source d'intérêt. Dans la section suivante, nous allons explorer de manière plus précise les ressources documentaires offertes par les folksonomies libres aux applications cartographiques à visée touristique.

3.3. Exploitation des services et expérimentation sur l'interopérabilité de sources ouvertes pour la cartographie touristique

Afin de poursuivre notre questionnement relatif d'abord à l'interopérabilité des bases de données ouvertes et à l'usage de données géographiques pour la production de cartes touristiques ensuite, nous procédons à une expérimentation. Cette expérimentation consiste en la réalisation d'une chaîne de production menant à la création d'une carte touristique. L'hypothèse testée est qu'en associant des sources de données compatibles, un nouveau point de vue surgit, différent de celui qui serait obtenu en se restreignant à une unique source d'informations. L'exercice est d'abord méthodologique. Nous nous demandons dans quelle mesure il est possible de proposer des recommandations touristiques, en mobilisant des sources de données institutionnelles et des folksonomies plus ou moins structurées. Nous avons proposé, au terme de la première partie, qu'OpenStreetMap puisse être une source unificatrice, offrant la possibilité de renseigner des informations locales et spécifiques. Cette proposition repose en partie sur l'idée selon laquelle les bases de données institutionnelles et les bases de données folksonomiques peuvent être complémentaires thématiquement, légalement et techniquement, afin de produire de nouvelles représentations cartographiques.

Dans le but de tester cette proposition, **nous réalisons une carte papier mettant en valeur des zones de la ville de Paris qui seraient attractives pour un tourisme urbain, tourné vers une promenade inspirée par la thématique du *Street Art* et par l'architecture**. Nous sélectionnons cinq sources de données : **OpenStreetMap, L'APUR, la Mairie de Paris, Wikidata et Flickr**[216].

De ces sources nous souhaitons, d'une part extraire des informations homogènes sur le territoire de Paris intra-muros et d'autre part, extraire des informations plus disparates décrivant des qualités originales de lieux qui les distingueraient d'autres lieux. En ce qui concerne ces informations disparates, nous considérons comme acceptable la possible absence d'exhaustivité. D'une part dans le cadre d'une réalisation industrielle nous procéderions à une campagne de contribution, à la manière des cartes régionales décrites en début du chapitre. D'autre part, la présence d'objets sera pour cette réalisation un facteur d'attractivité mais l'absence d'objet n'est pas un facteur de

[216] Le choix de cette thématique ne relève que d'un choix personnel d'appétence pour le Street Art mais il aurait été tout à fait possible de faire la même démonstration avec une tout autre thématique urbaine.

répulsion. Dans cet exemple, pour bénéficier des données institutionnelles exhaustives, nous sommes tenus de contraindre notre travail au périmètre dit *intra-muros* de la ville de Paris. Il aurait pu être intéressant de procéder au même processus en intégrant les communes limitrophes pour lesquelles une continuité piétonne existe avec Paris. Cependant, les données sélectionnées dans les ressources de la Mairie de Paris sont limitées au périmètre *intra-muros*.

La Mairie de Paris est un fournisseur majeur de données ouvertes, relatives au territoire qu'elle administre. Ce service s'organise autour du site Web https://opendata.paris.fr ouvert depuis janvier 2011, dont le catalogue de données est enrichi régulièrement. La page présentant le service Open Data Paris indique que le service considère les données en sa possession comme nécessaires à la réalisation de ses missions, mais également comme un patrimoine immatériel « qui peut être mis en valeur pour l'ensemble de la collectivité »[217]. Par collectivité, elle entend « les chercheurs qui peuvent y trouver matière à nourrir leurs travaux et expériences, les développeurs qui peuvent créer des services innovants utilisant ces données, les citoyens et journalistes qui y trouvent des informations brutes et les entreprises qui peuvent fournir une valeur ajoutée à ces données, et ainsi créer de l'emploi et de la richesse pour la collectivité ». Au travers de la démarche d'ouverture des données publiques, la Ville de Paris place « les habitants [comme] co-concepteurs des évolutions de leur Ville »[218].

L'Atelier parisien d'urbanisme (Apur), « a pour missions de documenter, analyser et développer des stratégies prospectives concernant les évolutions urbaines et sociétales à Paris et dans la Métropole du Grand Paris »[219]. À l'interface des acteurs de Paris et de la Métropole, l'Apur propose une expertise géographique relative aux politiques publiques d'aménagement et de développement aux échelles parisiennes et métropolitaines. Au 1er janvier 2019, on compte dans les statuts de l'Apur sept membres de droit (comme la Ville de Paris, l'État, la Métropole du Grand Paris et la RATP) et dix-neuf membres adhérents à l'association (comme le territoire Grand Paris Grand-Est, la Caisse des dépôts, Eau de Paris, Ile-de-France Mobilités et Paris Habitat). En

[217] Page intitulée « Paris ouvre ses données publiques », https://opendata.paris.fr/pages/lademarche/, consultée en mai 2019.
[218] Page intitulée « Paris ouvre ses données publiques », https://opendata.paris.fr/pages/lademarche/, consultée en mai 2019.
[219] Page de présentation de l'APUR : http://www.fnau.org/fr/agence/atelier-parisien-durbanisme/

2014, l'Apur commence son processus d'ouverture des données qu'elle possède par la mise en place d'un service en ligne et publie alors les limites administratives des communes, des départements et des EPCI, les limites administratives des quartiers et arrondissements parisiens, le dessin des plans d'eau et l'ensemble du réseau de transport en commun existant et en projet (hors bus). Aujourd'hui, l'atelier publie sous licence ODbL quatre-vingt-six jeux de données. Parmi ces jeux de données, citons celui des emprises bâties, des emprises des équipements comme les espaces verts, les terrains de sports et les hôpitaux ou encore les locaux commerciaux ayant une vitrine et un accès sur l'espace public.

Sur le territoire de la Ville de Paris, nous disposons donc de deux sources institutionnelles, Open Data Paris et l'Apur et de deux sources citoyennes, OpenStreetMap et Flickr. Les références entre les trois premiers producteurs existent : la documentation de l'Apur fait référence à la Ville comme une source de données et L'Apur et la Ville sont des sources de données pour les contributeurs à OpenStreetMap. En outre, Open Data Paris présente sur son site Web des représentations simples de ses jeux de données en superposition du fond de carte produit avec la base de données OpenStreetMap et proposé par Leaflet. Cette expérimentation met donc à l'épreuve l'utilisabilité conjointe de ces trois sources de données qui se connaissent et s'enrichissent mutuellement.

Lorsqu'on s'interroge sur l'utilisabilité des données publiques ouvertes, deux questions sous-jacentes apparaissent. D'une part, qui pourraient être les utilisateurs de ces données et d'autre part, de quelles ressources ont-ils besoin ? Deux chercheures en sciences de l'information et de la communication, Valérie Larroche et Valentyna Dymytrova (2018) apportent des éléments de réponse à ces questions. Elles recherchent, dans le cadre de l'ANR OpenSensingCity, les caractéristiques nécessaires d'une plateforme de partage de données ouvertes pour assurer la manipulation des données ouvertes dans des cadres de production déjà installés. Elles procèdent à une enquête auprès de professionnels utilisateurs de ces données. Ces professionnels sont identifiés comme appartenant aux catégories professionnelles de l'informatique et du traitement de données (« développeur, data scientists, data analysts ou data journalistes » (Larroche et Dymytrova, 2018)). La problématique de l'utilisation réelle des données ouvertes est celle de l'interopérabilité. L'interopérabilité doit être technique mais également sémantique. De ce fait, les

chercheures mènent des entretiens auprès des acteurs identifiés, afin d'évaluer dans quelle mesure l'exploitation de données ouvertes bénéficierait d'une intégration au Web sémantique. Au terme des entretiens, Larroche et Dymytrova (2018) distinguent quatre difficultés principales, rencontrées lors de l'exploitation des données ouvertes qui peuvent varier relativement à la fonction de l'enquêté (selon qu'il est responsable d'une infrastructure automatisée ou des aspects interprétatifs des données). Les données sont diffusées selon des formats disparates, elles ne présentent pas nécessairement une structure exploitable dans une chaîne de production déjà existante et d'autant moins lorsque la documentation est peu précise voire inexistante. Enfin, le procédé d'acquisition peut limiter les acquisitions de données lorsqu'une API n'est pas mise en place ou au contraire lorsqu'elle est trop complexe à mobiliser.

Face à ces difficultés, les attentes des professionnels interrogés consistent en une normalisation des données et des métadonnées. Un interrogé aborde l'émergence de l'*open data* local, au sein de certaines villes et souligne que celles-ci ne s'organisent pas à l'échelle nationale pour proposer des jeux de données interopérables techniquement et thématiquement. De ce fait, les chercheures posent la question de l'utilité des concepts du Web sémantique, déjà implémentés dans le Web actuel, comme « des solutions pertinentes à des problématiques liées à la réutilisation de l'Open Data » (Larroche et Dymytrova, 2018). Comme nous l'avons présenté dans le premier chapitre, ce Web sémantique est tourné vers la découverte et l'exploitation automatique d'informations. Ce type d'utilisation est cependant guidé par l'affectation de mots-clés par leurs producteurs selon une structure définie par le W3C. Une ressource du Web (une page HTML, une photographie, un objet d'une base de données comme OSM ou Wikidata) est alors porteuse d'une description lisible (dans le fichier textuel la décrivant) par un être humain et est, selon le principe du Web sémantique, normalisée de manière à être mobilisée par le moyen de procédés informatiques. Le Web sémantique repose sur des procédés techniques et méthodologiques que sont les modèles de données, les ontologies, le langage de requête SPARQL adapté au format standard RDF. Ces standards sont conçus par les nouveaux acteurs du Web et notamment par le W3C. Si Larroche et Dymytrova (2018) présentent un Web sémantique idéal qui apporterait davantage de signification aux données disparates de l'*open data*, elles nuancent cette promesse en mentionnant deux limites au processus d'intégration des données ouvertes au Web sémantique. D'une part, le Web sémantique

est une technologie complexe encore peu connue des experts enquêtés. D'autre part, se tourner vers cette technologie exige un investissement conséquent en formation et en force de travail. En effet, apporter la complexité nécessaire à l'utilisation étendue de données ouvertes ne peut reposer que sur l'implication d'utilisateurs humains. Larroche et Dymytrova (2018) mobilisent ici le concept de Web socio-sémantique ou Web cognitivement sémantique développé en 2005 par Zacklad, lorsqu'il le décrit comme visant « à soutenir les activités de recherche d'utilisateurs humains dans des corpus complexes et évolutifs » et l'opposant au Web sémantique qui « porte sur des contenus stables et assez simples pour être décrits en logique formelle ». L'approche par l'idée du Web socio-sémantique mène à la proposition de concevoir des ontologies locales liées pour faire face à l'hétérogénéité inhérente aux données ouvertes (Noyer et Carmes, 2012).

On retrouve là l'intérêt que nous avons exposé au premier chapitre, de mobiliser l'ontologie construite par les contributeurs à Wikidata afin de classer les données OpenStreetMap qui sont décrites selon un modèle très horizontal.

La quatrième source de données que nous mobilisons est la plateforme Flickr. Flickr est le nom d'un service de gestion et de partage de photographies en ligne. Anciennement propriété de Yahoo, il appartient, depuis avril 2018, au groupe SmugMug, spécialiste de la photographie publiée en ligne. De cette source, ce sont les métadonnées des photographies et non les images en elles-mêmes qui nous seront utiles. Ces métadonnées portent deux types d'informations qui, conjointement, vont nourrir notre classification des zones de promenades en ville, les coordonnées des photographies[220] et les *tags* qualifiant ces photographies écrits par leurs auteurs. Dans le premier chapitre, nous avons cité Flickr comme un cadre de production de folksonomie. Nous puisons donc, grâce à la folksonomie de Flickr, dans la subjectivité des contributeurs photographes pour mettre en évidence des lieux qui les ont inspirés. **Dans le cadre du prototype de carte touristique que nous construisons, nous exploitons également le caractère folksonomique du procédé de création de l'ontologie de Wikidata afin de sélectionner les photographies de la base de données Flickr.** En effet, une entrée thématique de notre carte consiste à guider les

[220] Nous considérons ces informations volontaires compte tenu du choix laissé au contributeur de rendre la position du lieu photographié visible ou non Il est possible que certains utilisateurs n'aient pas conscience que leurs photographies portent un *geotag* qui est inscrit automatiquement par certains appareils.

touristes vers l'expression artistique qui s'intègre dans les rues de la ville et orne les façades des immeubles de l'espace public. L'appellation générale que nous retenons est celle de *Street Art*. Il nous a semblé qu'un domaine d'expression aussi dynamique pouvait répondre à plusieurs appellations faisant référence à différents courants. Wikidata est donc, dans cet exemple, le réceptacle de ces appellations, nous permettant d'interroger des ressources du Web sémantique sur l'ensemble du champ-lexical du *Street Art*. La base de données Wikidata pourrait être une source d'objets géographiques pour notre expérimentation. Néanmoins, il nous semble que ces informations complémentaires seraient davantage exploitables pour la production d'une carte numérique. En effet, les informations disponibles dans Wikidata peuvent être spécifiques à chaque objet et non pérennes. En conséquence, elles n'auraient pas leur place dans OSM. Par exemple, le nom de l'artiste d'une œuvre de *Street Art*, le nombre de concerts et les noms des artistes qui se seraient produits sur l'esplanade de l'Hôtel de Ville ou encore le code ISBN de l'œuvre *Notre-Dame de Paris*, associés à l'objet décrivant la Cathédrale Notre-Dame sont autant d'informations, qui ne sont pas représentables sur une carte papier mais qui peuvent être figurées à la demande de l'utilisateur d'une carte numérique interactive.

3.3.1. Construction d'une base de données spatiales

À partir de ces quatre sources disparates, nous constituons un agrégat de données en utilisant le système de gestion de base de données PostgreSQL /PostGIS. Nous nous positionnons ainsi dans la perspective de pouvoir rendre publiques les productions qui seront issues de cette base de données. La publication implique ainsi l'application des licences associées aux données.

La première étape de la constitution de la base de données consiste à assurer une compatibilité entre les licences ainsi qu'une compatibilité des licences avec les procédés de transformation que nous mettons en place.

Durant un processus de création, tel que celui que nous présentons, des sources de données ou des éléments d'une source de données peuvent être exclus, lorsque les restrictions imposées par leur licence contraignent trop fortement le processus de création. Par exemple, nous avons envisagé de ne pas conserver le jeu de données

composé des métadonnées relatives aux photographies de la base de données Flickr. La liberté de description permise aux contributeurs à Flickr est similaire à celle permise aux contributeurs à OSM. En fait, le contributeur à Flickr est davantage libre dans sa contribution que celui d'OSM. Le choix de la licence de la photographie revient au contributeur à Flickr[221]. Au contraire, en s'inscrivant à openstreemap.org, les contributeurs à OSM acceptent que leurs contributions intègrent une base de données unique et donc protégée d'une unique licence qui lui est propre. La licence ODbL prévoit l'attribution du producteur de données sous la forme suivante : « Vous mentionnez la paternité : Vous devez mentionner la source de la base de données pour toute utilisation publique de la base de données, ou pour toute création produite à partir de la base de données, de la manière indiquée dans l'ODbL. Pour toute utilisation ou redistribution de la base de données, ou création produite à partir de cette base de données, vous devez clairement mentionner aux tiers la licence de la base de données et garder intacte toute mention légale sur la base de données originaire. La source devra être mentionnée de la façon suivante : « [Nom de la base de donnée], [Producteur de la base de donnée], [date], sous licence ODbL »[222]. La Mairie de Paris propose un exemple de bonne pratique pour créditer les données qu'elle publie sous licence ODbL : « Éclairage public - Mairie de Paris, 23/02/2018, sous licence ODbL ». Appliquée au producteurs que constituent les contributeurs à OSM, la fondation OSM propose de citer ainsi la source : « © OpenStreetMap contributors ».

Au contraire de la protection légale unifiée dans le cadre d'OSM, onze choix de licences sont offerts par Flickr. Finalement, pour 80 522 photos localisées à Paris du 23 mars 2018 au 23 mars 2019, seules 330 sont protégées d'une licence rigoureusement compatible avec l'ODbL (Les licences : « Aucune restriction de copyright », « Travaux du gouvernement américain », « Domaine public (CC0) » et « Marque du Domaine Public ». La licence CC-BY 2.0 choisie par Flickr est à la limite entre la compatibilité des licences libres et l'incompatibilité entre les mécanismes d'attribution (fig. 53). En effet, la licence Creative Commons Attribution version 2 (CC-BY 2) décrit l'attribution ainsi : « Vous devez créditer l'Œuvre, intégrer un lien vers la licence et indiquer si des modifications ont été effectuées à l'Œuvre. Vous devez indiquer ces informations par tous les moyens raisonnables, sans toutefois suggérer que l'Offrant vous soutient ou

[221] Nous considérons, que la licence choisie pour une photographie enregistrée sur Flickr s'applique également aux métadonnées qui lui sont associées.
[222] Traduction réalisée par l'association Veni, Vidi, Libri et réutilisée par la Mairie de Paris.

soutient la façon dont vous avez utilisé son Œuvre ». Le principe à l'origine de cette condition est le même que celui pour l'ODbL. Il n'existe dans ces deux clauses qu'une simple demande : l'auteur du document utilisé doit être cité. Cependant, en ce qui concerne l'ODbL, le document soumis à la licence est la base de données. Dans le cas d'OSM, les contributeurs se déclarent, à leur inscription, comme faisant partie d'un groupe que l'on reconnaît comme le producteur. Dans le cas des photographies contenues dans Flickr, chaque photographie porte une licence qui lui est dédiée et un auteur individuel. L'attribution des photographies issues de Flickr doit renseigner le nom de tous les photographes auteurs des photographies utilisées. Il est précisé que le crédit doit être indiqué « par tous les moyens raisonnables ». Lorsqu'il s'agit d'une carte imprimée, il semble impossible d'inscrire l'ensemble des noms des photographes dans l'espace généralement réservé aux sources. Une solution envisageable consiste en une liste détaillée intégrée dans une page Web. Dans le cadre de la construction d'un prototype, cette solution n'est pas envisagée. Nous proposons cette liste en annexe 1. Ainsi, nous proposons un système d'attribution adapté au format de la thèse et respectueux de la licence CC-BY et nous ajoutons 2743 métadonnées au 330 déjà retenus. Si le jeu de données est réduit de 96%, le nombre de photographies attestant la présence d'une œuvre *Street Art* reste conséquent, nous avons donc choisi de conserver cette source de données.

Identifiants et intitulés des licences dans l'usage de Flickr	Compatibilité avec la licence ODbL	Nombre de photographies
0, Tous les droits sont réservés.	Licence fermée. Incompatible.	64838
1, Attribution - usage non commercial - partage à l'identique (CC-BY-NC-SA).	Licence de libre diffusion. Incompatible car elle restreint à un usage non commercial et exige un partage à l'identique, c'est-à-dire dans ce cas sans usage commercial, des œuvres produites à partir du document.	6572
2, Attribution - usage non commercial (CC-BY-NC).	Licence de libre diffusion. Incompatible car elle restreint à un usage non commercial.	1939
3, Attribution - usage non commercial - pas de dérivés (CC-BY-NC-ND).	Licence de libre diffusion. Incompatible car elle restreint à un usage non commercial et à la modification.	2007
4, Attribution (CC-BY).	Licence libre non *copyleft*. Compatible.	2743

5, Attribution et partage à l'identique (CC-BY-SA).	Licence libre *copyleft.* Incompatible, le partage à l'identique de cette licence ne se propage pas uniquement aux données utilisées conjointement mais aux œuvres produites à partir de ces données. Cette clause est contradictoire avec la notion de « produced work » de la licence ODbL.	1790
6, Attribution - pas de dérivés (CC-BY-ND).	Licence de libre diffusion. Incompatible car elle restreint à un usage sans modification possible.	335
7, Aucune restriction de copyright	Licence libre non *copyleft*. Compatible.	0
8, Travaux du gouvernement américain	Licence libre non *copyleft*. Compatible.	0
9, Domaine public (CC0)	Licence libre non *copyleft.* Compatible.	66
10, Marque du Domaine Public	Licence libre non *copyleft*. Compatible.	262

Figure 53 Compatibilité des licences Flickr avec la licence ODbL (Hayat, 2019)

Une fois que nous nous sommes assurés de la cohérence légale entre les sources de données, nous procédons à leur obtention et à leur chargement selon un format technique de base de données relationnelle (PostgreSQL/PostGIS).

La deuxième étape de cette création consiste en la sélection, au sein de ces sources disparates, d'informations en adéquation avec notre recherche de zones propices à la promenade urbaine. L'agrégat de tables spatiales présente alors dix tables de données partageant la même projection adaptée à celle de Paris, la projection Lambert 93 (figure 54). Nous avons présélectionné les objets des métadonnées Flickr dont la licence est compatible à l'ODbL et qui répondent au champ lexical du *Street art*. Nous créons sept tables d'objets issus de la base de données OSM. Et nous chargeons les deux jeux de données sélectionnés de l'APUR et de la Marie de Paris.

Jeu de données	Source	Licence
Paris intra-muros, limite administrative.	OpenStreetMap (25/03/2019).	ODbL
Cours d'eau.	OpenStreetMap (25/03/2019).	ODbL
Bâtiments classés dans un référentiel patrimonial.	OpenStreetMap (25/03/2019).	ODbL
Arrêts de bus.	OpenStreetMap (25/03/2019).	ODbL
Réseau routier.	OpenStreetMap (25/03/2019).	ODbL
Points de vue.	OpenStreetMap (25/03/2019).	ODbL
Restaurants, cafés et jardins offrant l'accès à la Wifi.	OpenStreetMap (25/03/2019).	ODbL
Métadonnées des photographies géotagées de la base de données Flickr.	Flickr (23/03/2018 au 23/03/2019).	CC-BY 2 ou sans *copyright* ou Domaine Public
Les espaces de terrasses parisiens.	Open Data Paris (créé le 28/12/2016, mis à jour au 23/03/2019).	ODbL
Emprises bâtiments parisiens.	APUR (créé en 2014, mis à jour au 15/02/2019).	ODbL

Figure 54 Sélection de dix jeux de données parmi les quatre sources données aux licences compatibles (Hayat, 2019)

3.3.1.1 Extraction, sélection et analyse des données de Flickr

Grâce à l'API Flickr, nous procédons à l'extraction de métadonnées de photographies géotaguées et postées du 23 mars 2018 au 23 mars 2019. Au travers de ce jeu de données, nous cherchons la notification par des promeneurs-photographes de traces de *Street Art* dans les rues de Paris. Il est évident que les 80 000 photographies extraites, dont nous ne conservons qu'une part très réduite, ne recensent pas l'ensemble des *graffitis*, pochoirs et installations de rues, qui ont pu être exposés entre mars 2018 et mars 2019. Cependant, nous considérons avoir retenu des localisations significatives où des artistes, peintres et photographes ont choisi de s'exprimer. Ce jeu de données porte l'expression de l'engouement de chaque photographe pour un ou plusieurs lieux. En outre, nous exploitons la folksonomie construite par les photographes-contributeurs

afin de retrouver ces lieux d'expressions de Street Art. En notifiant un *tag* qui relève de ce courant artistique, nous sommes assurés que la photographie vise à valoriser cet élément précis.

Afin de collecter l'essentiel des métadonnées des photographies relatives au *Street Art*, nous utilisons la base de données de connaissance Wikidata pour définir un champ lexical autour de ce terme (fig. 55). Nous exploitons la folksonomie précisément réglementée des contributeurs à Wikidata, afin de sélectionner des objets classés selon une folksonomie plus personnelle, celle des contributeurs à Flickr.

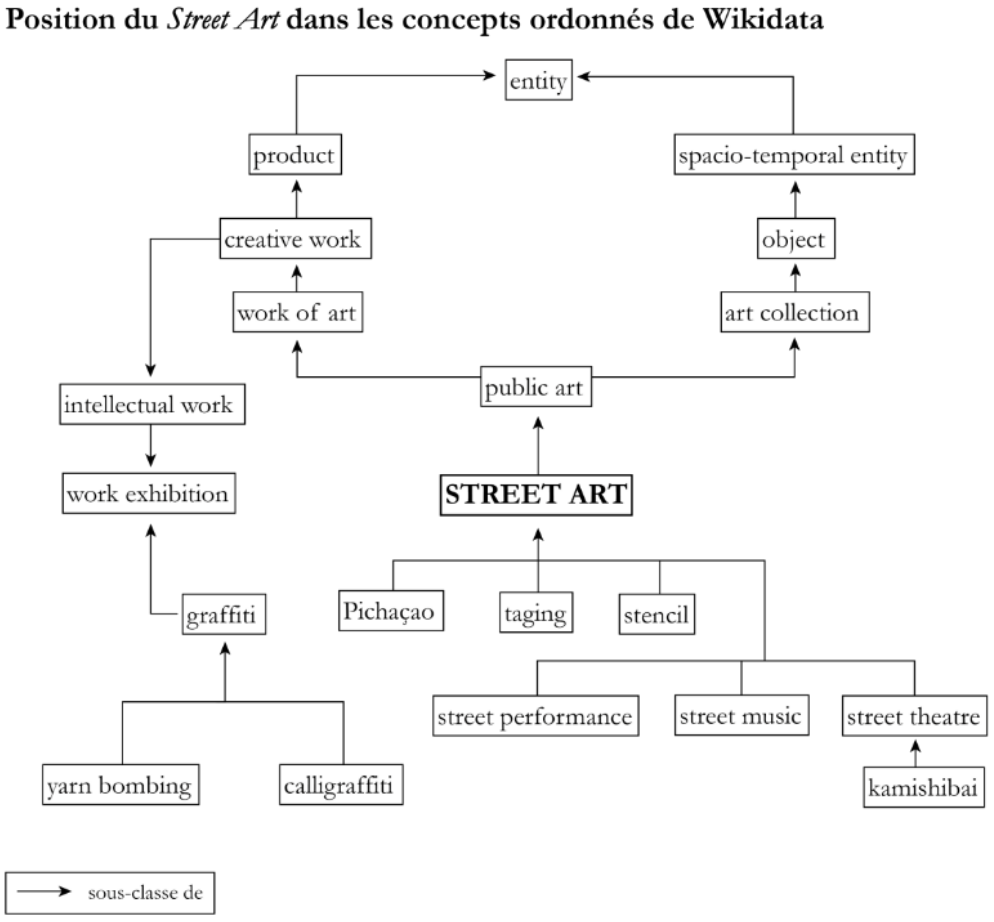

Figure 55 Exploitation de la classification du Street Art et de ses sous classes afin de construire un champ lexical associé (Hayat, 2019)

3.3.3.2. Extraction, sélection et analyse des données d'OSM

La base de données OpenStreetMap apporte à ce prototype deux perspectives. D'une part, à la manière de l'exploitation de la base Flickr, nous mobilisons l'intérêt des contributeurs pour certaines thématiques originales. D'autre part, nous nous appuyons

sur des données que nous estimons, pour notre usage, suffisamment riches en nombre et en attributs pour constituer des données comparables sur l'ensemble du périmètre.

En ce qui concerne les informations originales, nous avons retenu **les restaurants, cafés et jardins offrant l'accès à la Wifi, les points de vue et les bâtiments classés dans un référentiel patrimonial.** Ces trois informations ne sont pas exhaustives dans la base de données. Cependant, dans une perspective touristique, nous considérons que leurs renseignements assurent l'existence du service ou du lieu d'intérêt. Un point de vue est, le plus souvent, renseigné lorsqu'un panneau le signale mais nous présumons les contributeurs capables d'estimer l'attractivité d'un lieu pour la perspective qu'il offre à voir. Afin de collecter les bâtiments patrimoniaux, nous sélectionnons les objets portant le *tag* `heritage=*`. Les classements patrimoniaux nationaux et internationaux sont rendus publics régulièrement. Le projet OSM se positionnant dans une perspective d'objectivation considère les lieux patrimoniaux uniquement par le prisme des listes patrimoniales. Nous exploitons ces classifications davantage comme la reconnaissance par les sociétés d'une originalité du lieu classé. Cependant, le renseignement de cette classification reste incomplet dans OSM. Nous notons la complexité de la documentation relative à la clé `heritage` (Key:heritage, 2019) sensée proposée une modélisation normalisée pour intégrer les listes patrimoniales nationales et internationale. En effet, la modélisation du référencement du patrimoine mondial doit prendre en compte son caractère réticulaire et les controverses politiques qui accompagnent la préservation du patrimoine (Severo, 2014). Ceci affecte faiblement notre propre classification car les zones sont valorisées par la présence de lieux patrimoniaux et ne sont pas déclassées par leur absence. Enfin, l'accès au wifi, désigne un service largement utilisé pendant une promenade lorsque le promeneur est en quête d'informations plus détaillées sur son environnement. Toujours dans une perspective touristique, nous pourrions imaginer des renvois, inscrits dans l'habillage de la carte, vers des ressources en ligne.

Nous considérons le jeu de données du réseau routier parisien à la fois exhaustif et original. Pour affirmer le caractère exhaustif du réseau parisien, nous ne procédons à aucune évaluation qualitative de comparaison à une référence. C'est par la pratique régulière de ce jeu de données que nous sommes en mesure de l'affirmer. En outre, nous ne sommes pas à la recherche de voies de service confidentielles ou d'impasses

privées mais de rues principales et piétonnes de la ville. Nous considérons donc le jeu de données complet en termes de voies[223]. Cependant, selon l'usage, la qualité du réseau routier OSM peut être insatisfaisante. Si on considère le jeu de données dans le but de réaliser la cartographie routière de Paris, une campagne de contribution est nécessaire. En effet, la continuité du réseau n'est pas garantie. On relève souvent des changements de types de classification inadaptés à la cartographie. Par ailleurs, même lorsque la continuité entre les types de voies est conservée, on juge parfois une voie classée trop haut dans la hiérarchie et inversement. Ceci s'explique par des écarts d'interprétation dans la définition de ce qu'est une voie primaire par rapport à une voie secondaire. D'ailleurs, le débat n'est pas clos entre les contributeurs sur la question de savoir si le réseau routier doit être hiérarchisé selon des critères mondiaux ou locaux. Un réseau routier conçu selon des critères mondialement homogènes permettrait une exploitation automatique et la création de cartes numériques mondiales multi-échelles, sans avoir à considérer des spécificités selon les zones géographiques. Pour ce qui est d'une cartographie à grande échelle pour une publication papier sans objectif de reproductibilité sur un autre périmètre, cette disparité essentiellement mondiale ne nous concerne pas. Si cette situation pose une réelle difficulté pour la production de cartes routières, il s'agit dans notre cas d'attributs utiles à la classification des zones attractives. En effet, la prise en compte de la forme locale du réseau intègre une subjectivité qui, du point de vue du cartographe, apporte de la thématique à l'exploitation de la carte. La figure 56 en donne un exemple.

[223] Cette densité nécessite pour une cartographie lisible que le réseau routier soit généralisé, même à grande échelle.

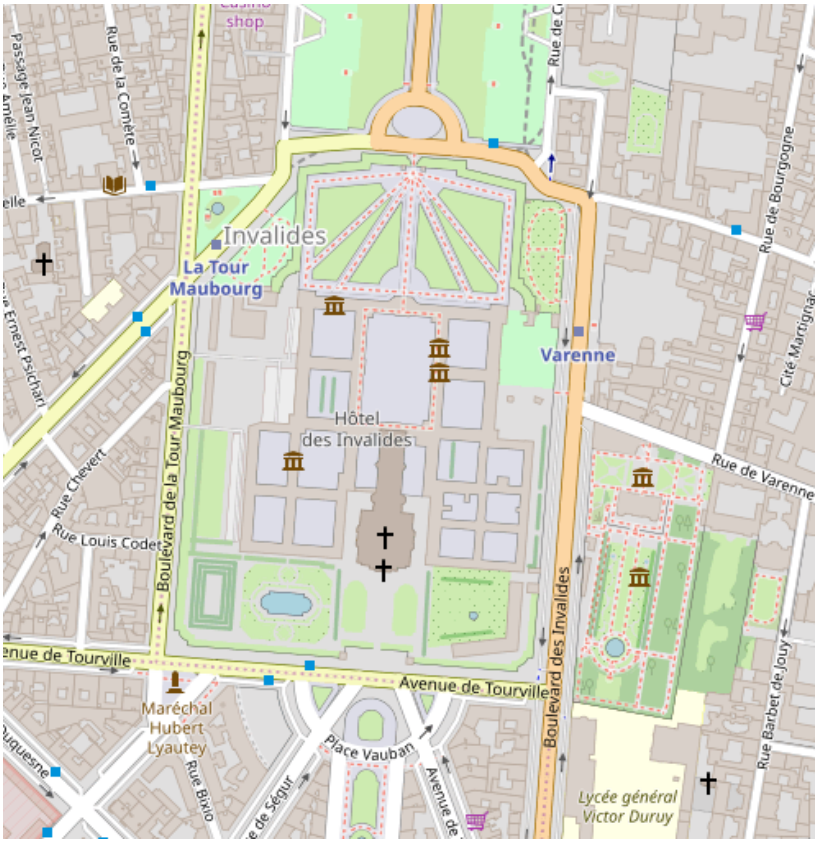

Figure 56 Discontinuités de classification du réseau (capture d'écran de la slippy map *d'openstreetmap.org)*

Enfin, nous retenons trois jeux de données utiles à notre projet mais qui ne présentent aucune originalité propre à la liberté de contribution. Nous avons extrait la limite administrative de la Commune de Paris, le contour des espaces en eau et les arrêts de bus. La limite administrative de Paris est également disponible sur le site d'Open Data Paris, le choix de la source OSM repose sur la commodité technique d'utiliser les données OSM toutes chargées en base de données en local sur notre ordinateur. Les surfaces en eau ont été filtrées afin de ne conserver que l'emprise de la Seine et du canal Saint-Martin, le bassin de la Villette et le canal de l'Ourcq. Les arrêts de bus parisiens nous apportent l'information du niveau d'accessibilité des zones, par un moyen de transport adapté à la poursuite de l'exploration du paysage urbain. Deux sources de données étaient à envisager : le référentiel des arrêts de transports en commun en Ile-de-France publié par *Île de France Mobilités*, l'organisme responsable des transports en commun en Île-de-France ou OpenStreetMap. Si *Île de France Mobilité* est l'institution de référence en terme d'infrastructure de transports, il est nécessaire, afin de mesurer un certain niveau d'accessibilité, de disposer d'arrêts de bus produits en

cohérence avec le réseau routier dont nous disposions. En effet, une étude réalisée par l'association Jungle Bus, nous apprend que si le jeu de données officiel présente une exhaustivité plus importante dans le périmètre de Paris intra-muros qu'OSM (21% d'arrêts de bus supplémentaire), il contient certaines erreurs de position notables (vérifié sur le terrain ou à l'aide de l'aide d'imageries). L'activité dynamique de cette association de contributeurs qui se consacrent à cette thématique nous a guidés vers la sélection de la production d'OSM plutôt que vers celle d'*Ile de France Mobilité.*

3.3.3.3. Extraction et sélection des données de la Mairie de Paris et de l'APUR

Les données de l'APUR et de la Mairie de Paris sont deux jeux de données institutionnels. Nous insistons sur ce qualificatif car leur rôle de producteurs de données au service des politiques publiques motive la production des jeux de données que nous mobilisons.

C'est ce rôle d'institution qui permet à la Mairie de Paris de recenser l'installation des étalages et terrasses sur la voie publique. La délimitation précise de ces installations est relevée régulièrement et le jeu de données est donc constitué d'objets de type polygones. Les attributs notables qui s'y rapportent sont le type de commerce dont il s'agit (restaurant, boulangerie, bar, brasserie) et la période d'ouverture de la terrasse. L'Apur, n'a pas, à la différence de la Mairie de Paris, de rôle de contrôle du bon usage de l'espace public. Cependant, son statut d'institution lui donne accès aux ressources de deux institutions publiques, la DGFiP et la Mairie de Paris. Comme on peut le lire dans la documentation, l'Apur s'appuie sur les services de l'entreprise Interatlas. C'est sur ces trois ressources, que, comme cela est expliqué dans la documentation accompagnant le jeu de données sur les emprises bâties à Paris, l'Apur a constitué « une couche de départ la plus complète et la plus à jour possible sur laquelle les mises à jour et les corrections pourraient ensuite être effectuées ». L'APUR s'appuie également sur les ressources publiques comme le Fond Topographique Cadastral ainsi que sur le volume bâti, mesuré par la direction de l'Urbanisme de la Ville de Paris et la BD Topo de l'IGN. À ces ressources obtenues par partenariats, l'Apur y ajoute des données qu'elle produit comme la date des bâtiments, leur hauteur ou encore le type de toiture.

3.3.3.4. Une base de données géographique ?

Une des premières questions à résoudre est celle de l'agrégation de ces données disparates : dans quelle mesure peuvent-elles constituer une base de données relationnelle et géographique fiable ?

Dans le premier chapitre, nous avons défini les bases de données géographiques comme le résultat d'une représentation sélective et méthodique du monde réel. Nous considérons que l'échelle selon laquelle les objets sont simplifiés, pour procéder de cette transformation de la réalité à la donnée, doit être, autant que possible, homogène. Nous avons également distingué la base de données spatiale de la base de données géographique selon la place accordée à la réflexion géographique qui précéderait la construction de la base de données.

Il apparaît rapidement que ce prototype n'est pas celui d'une base de données relationnelle au sens strict méthodologique. Les objets que nous manipulons ne sont que rarement liés par des relations inscrites dans leurs attributs qui seraient pensées au service d'une interopérabilité technique et conceptuel. Ces rares liens sont néanmoins des ébauches de possibles futures bases de données, précisément géographiques. Dans le premier chapitre, nous avons donné l'exemple d'une relation qui consiste à affecter à des objets classés comme des villes, l'attribut les définissant comme le centre administratif d'un territoire spécifique lui-même renseigné dans la base de données. La relation lie le territoire et la ville, deux types d'objets autonomes dans leurs descriptions et qui peuvent être mobilisés ensemble grâce à la modélisation de leur relation. La modélisation de cette relation prévoit des contraintes comme par exemple l'exigence que le territoire et la ville appartiennent eux-mêmes à un objet englobant comme celui décrivant le pays. Au sein de l'agrégat que nous avons constitué, aucune relation n'existe initialement entre les quatre sources que nous mobilisons.

Il paraît donc excessif, à ce stade, de qualifier notre ensemble de données de base de données. Nous nous demandons néanmoins s'il spatial ou géographique ? Les contextes de production que sont OpenStreetMap, l'Atelier parisien d'urbanisme (Apur) et la Mairie de Paris sont axés sur la modélisation exhaustive et thématiquement large de territoires. D'ailleurs, dans ce souci, l'Apur précise, dans la documentation du jeu de données des emprises bâties, les usages pour lesquels le jeu de données peut être utilisé. C'est l'engagement d'une certaine qualité des données qui est inscrit. L'Apur considère

que l'usage statistique ou de représentation globale du bâti est compatible avec la qualité du jeu de données. Au contraire, l'usage ponctuel n'est pas recommandé. Il est admis par ces producteurs que les objets sont conçus conjointement et notamment d'un point de vue topologique.

Il nous est impossible de connaître complètement le procédé de modélisation des bases de données de l'Apur ou de la Mairie de Paris qui ne présentent dans les métadonnées que le résultat de leur modélisation et non son processus de création. Néanmoins, nous décelons des indices de conception géographique initiale à la création des bases de données lorsque nous constatons des choix géométriques conventionnels (comme les contraintes topologiques et géographiques détaillées par l'Apur dans ses documentations), la sélection de la projection Lambert 93 ou encore le détail de procédures scientifiques qui ont mené par exemple, à la redéfinition, en 2015, des zones touristiques internationales à Paris[224]. Nous avons identifié des interactions entre la Marie de Paris et l'Apur au travers de leurs hyperliens de référence, des usages mutuels de leurs données et même de références communes, comme le code des voies défini par la Ville de Paris et réutilisé par l'Apur. En outre, si nous avons observé des décalages géométriques et des erreurs topologiques mineurs entre les objets du jeu de données des terrasses de la ville de Paris et celui des bâtiments de l'Apur nous pouvons supposer qu'il existe une mise en commun lors de la construction des jeux de données entre deux fournisseurs de données partenaires.

En ce qui concerne OpenStreetMap, nous avons qualifié, dans le premier chapitre, son contexte de production comme géographique. Cette qualification correspond à la finesse de description du contenu de certaines pages du *wiki* comme celles qui définissent les bonnes pratiques de description (et notamment l'exemple des rives d'un cours d'eau), le rappel constant de nécessité du respect de la topologie pour le réseau routier, les cours d'eau, les réseaux de transports ou encore les typologies d'usage des sols largement débattues. Ainsi, les données de l'Apur, de la Mairie de Paris et d'OSM sont conçues d'une telle façon, dans le respect de contraintes du domaine de la géographie, que des relations pourraient techniquement être reconstruites. En outre, ces productions pourraient intégrer, à l'avenir, dans leur modélisation des relations

[224] Documentation accessibles dans la rubrique travaux du site de l'APUR et également disponible sur le site d'Open Data Paris : https://www.apur.org/fr/nos-travaux/zones-touristiques-internationales-paris-zti-diagnostic-initial-commerces, consulté en mai 2019.

permettant d'associer les objets se rapportant aux mêmes concepts du fait de leur conception similaire.

À l'inverse, les métadonnées des photos Flickr sont l'exemple typique des données spatiales. La géolocalisation des photographies est l'unique description relative à l'espace qui soit partagé par tous les objets du jeu de données que nous avons constitué. Nous pouvons retrouver la mention de toponymes comme « 9ème arrondissement », « Étoile », « Jardin des plantes » et « Montparnasse », mais il s'agit de choix individuels, parfois partagés mais largement personnels. De nouveau, cet agrégat de jeux de données ne répond que partiellement à notre définition d'informations géographiques.

Nous ne nous limitons cependant pas à cet état intermédiaire qui consisterait à superposer graphiquement un ensemble d'objets. En effet, on procède à une sélection guidée par une question géographique et l'identification d'éléments de réponses adaptées au territoire. Notre idée est de proposer une représentation de Paris, la découpant en zones plus ou moins propices à la découverte des originalités urbanistiques de la ville.

La troisième étape de création consiste, alors, en la construction de nouvelles informations géographiques issues de l'agrégat de données. Afin de répondre à notre objectif de typologies, et en ne s'imposant aucun critère d'implantation et forme attendue pour le « zonage », trois approches de la question ont été envisagées :

1. À l'échelle des rues, nous pourrions classer les rues selon les *tags* qu'elles évoquent aux photographes-contributeurs (Flickr), selon la hauteur moyenne des constructions (APUR), selon la densité de surface de terrasses (Open Data Paris) et selon la présence d'ateliers d'artistes locaux (OSM). La représentation serait alors porter par des objets linéaires.

2. Une autre perspective géographique consisterait à reprendre des critères similaires, tout en suivant la continuité piétonne des berges de la Seine et du Canal de l'Ourcq ; ou dans la même perspective, en suivant la succession de boulevards au nord de Paris rejoignant la place des Ternes au quartier de la Villette : boulevard de Courcelles, boulevard des Batignolles, boulevard de Clichy, boulevard de Rochechouart, boulevard de la Chapelle et enfin le boulevard de la Villette. Plutôt

qu'une représentation de l'ensemble de la ville, nous nous concentrerions sur les alentours de ces axes majeurs de la ville.

3. Finalement, nous avons choisi une troisième perspective en considérant l'ensemble de Paris intra-muros. Deux découpages de la ville ont été envisagés. Le premier consistait à découper la ville en fonction des voies de circulation les plus importantes. La ville est quadrillée de grands boulevards circulaires et de voies larges intermédiaires. Ces voies principales, souvent encombrées, pourraient être considérées comme trop bruyantes pour être attractives. En les utilisant pour définir des zones dans Paris, nous les intégrerions à la représentation de la ville tout en proposant des zones relativement isolées des carrefours encombrés.

   Le second zonage envisagé, finalement retenu, consiste en un quadrillage régulier de la ville en cellule de 500 mètres de côté. La surface des cellules homogènes nous permet d'éviter un effet de taille dans la mesure de nos paramètres d'attractivité.

Ainsi, ce sont les cellules de 500 mètres de côté qui porteront la nouvelle information géographique que nous produisons dans cet exemple.

3.3.2. Des critères croisés

La carte que nous cherchons à produire doit mettre en évidence les quartiers[225] qui correspondent le plus aux critères définis en amont de la conception de la carte et que nous détaillons ci-dessous. Les critères considérés sont uniquement mélioratifs. Dans une version plus étoffée du prototype, il serait envisageable d'ajouter quelques critères répulsifs en rapport avec le sujet de la carte. Nous sommes à la recherche de zones de promenades, où trouver à la fois des espaces de repos comme les jardins et les parcs ou encore les terrasses de café et à la fois des formes artistiques dans le paysage urbain comme des bâtiments patrimoniaux ou des œuvres de *Street Art*.

[225] Nous employons le terme de quartier non pas en son sens administratif de la ville parisienne mais pour indiquer que les cellules ne sont pas considérées individuellement. Chaque cellule doit être envisagée comme un potentiel point de départ de promenades. De ce fait, une cellule présentant un rang élevé et qui côtoierait des cellules moins bien classées doit être considérée comme potentiellement à l'interface de cellules également attractives. Elles constituent un périmètre, ou quartier, plus attractif que les 250 000m² mis en évidence par la cellule manifestement attractive.

Nous avons retenu neuf critères pour leur pertinence thématique comme pour leur originalité et la diversité de leurs producteurs :

- *Proximité de La Seine et des canaux*

À l'échelle de chaque cellule, nous considérons que la présence d'un objet faisant partie de la description de la Seine, du canal Saint-Martin, du bassin de la Villette ou du canal de l'Ourcq est attrayante. Nous ne considérons pas la surface de la cellule occupée par une de ces surfaces en eau. En effet, la taille réduite des cellules (500 mètres de côté), laisse supposer que les utilisateurs de la carte sont amenés à sortir des limites des cellules. C'est donc la proximité immédiate aux quais et aux perspectives offertes qui sont valorisées. Si une cellule intersecte un de ces espaces en eau, nous affectons un point supplémentaire à cellule. La présence de ces surfaces en eau est représentée par l'image F dans la figure 59 (page 278). En outre, les cellules entièrement occupées par une surface en eau ne sont pas exclues du traitement. Celles-ci obtiennent un point.

- *Street Art*

La géolocalisation des photos produites à destination d'une publication sur Flickr présente des niveaux de qualité qui dépendent du matériel dont dispose le contributeur. Nous ne pouvons présumer d'une qualité au mètre : elle se rapproche plus probablement d'une qualité à 10 mètres, notamment dans en zone urbaine. En outre, les photographies peuvent être prises selon différents points de vue : gros plan, plan moyen ou plan large. De ce fait, nous ne pouvons pas déterminer sur quel bâtiment a été réalisé l'œuvre. Cependant, comme pour le critère de proximité aux cours d'eau, la taille réduite des cellules nous autorise à ne considérer que la présence d'une photographie étiquetée d'un mot du champ lexical du *Street Art* pour comptabiliser un point à la cellule. La présence d'au moins une photographie dans une cellule est représentée par l'image E dans la figure 59 (page 278). Le centre-est parisien et le nord du 13ème arrondissement se distingue particulièrement par la présence continue de photographies. Les quartiers périphériques nord et sud ne sont que rarement marqués par la présence d'une photographie *taggés Street Art*.

- *Rues piétonnes*

Espaces de déambulation, la valorisation des rues piétonnes est la solution sélectionnée afin de prioriser les zones moins denses en véhicules motorisés par rapport aux zones à

fortes circulations que sont les boulevards de la ville. Pour chaque cellule, nous calculons, en pourcentages, la part représentée (en mètres) par les rues piétonnes de la cellule sur l'ensemble des routes de la cellule (en mètres). Les rues piétonnes sont identifiées par les *tags* `highway=pedestrian` et `highway=living_street`. L'ensemble des routes est composé des tronçons portant la clé `highway`. L'image H de la figure 60 (page 279) met en évidence une forte densité de rues piétonnes au centre de la ville, dans les premiers arrondissements et une surreprésentation de ces voies à l'est de la ville par rapport à l'ouest.

- *Espaces de promenade*

Les espaces de promenade sont à la fois des zones piétonnes, comme des places, et des espaces verts. Pour la première catégorie, les objets retenus sont des chemins fermés qui portent les *tags* `highway=pedestrian` et `highway=living_street`. La seconde catégorie fait appel à deux clés, qualifiant également des chemins fermés. Les *tags* sont les suivants : `leisure=garden, leisure=park` et `natural=wood`. La surface que représente chacun de ces objets par cellule est rapportée à la surface totale des cellules à laquelle on soustrait la surface totale des bâtiments renseignés dans OSM. Nous mobilisons la couche d'information des bâtiments issue d'OSM afin d'exploiter les relations topologiques qu'ils partagent avec les espaces de promenade construits à partir des données OSM. Le résultat final est un pourcentage. L'image A de la figure 58 (page 277) présente la ville séparée en deux parties par une diagonale partant de l'ouest parisien jusqu'au sud-est. Les deux bois que sont Boulogne (à l'ouest) et Vincennes (à l'est) en sont les extrémités. Le sud parisien dispose de davantage de jardins et de places piétonnes.

- Bâtiment patrimonial

Les bâtiments patrimoniaux répondent à la volonté de signaler des espaces de la ville marqués par une richesse patrimoniale. Notre sélection s'est portée sur l'ensemble des chemins et des relations portant la clé `heritage`. Dans OSM, la clé `heritage` indique le classement d'une zone, d'un bâtiment, d'une œuvre pérenne dans un recensement patrimonial. Il s'agit d'objets géographiques qui bénéficient d'une protection juridique et technique adapté à l'ambition de leur conservation. Nous rapportons la surface des objets portant la clé `heritage` de chaque cellule à la surface

de l'ensemble des bâtiments issus d'OSM de la même cellule. L'image B de la figure 58 (page 277) présente une surreprésentation côté rive droite de la Seine des bâtiments patrimoniaux. Les quartiers centraux sont également plus largement couverts en bâtiments recensés dans les projets de préservation du patrimoine historique.

- Accessibilité au réseau Internet

Le critère relatif à l'accès au réseau Internet par un service de Wifi public nous permet d'ancrer la pratique de la promenade dans un possible contexte de contribution. Nous imaginons un accompagnement à la promenade urbaine largement construit à partir de données contributives. Il est peut-être possible d'encourager l'utilisateur à être lui-même contributeur. L'accès à la Wifi est un service qui limite le coût de la contribution du point de vue de la consommation des données mobiles comprises dans l'abonnement téléphonique de l'utilisateur. Dans une cellule de 500 mètres de côté, nous considérons que l'accès à un service de wifi est suffisant. La présence d'un seul objet portant le *tag* « `internet_acces=wlan` » apporte un point supplémentaire au rang de la cellule évaluée. D'après l'image D de la figure 59 (page 279), dans un rayon de 500 mètres la quasi-totalité de la surface de la ville est couverte par au moins une borne Wifi accessible au public. Les bois périphériques ne présentent pas ce service.

- Hauteur moyenne des bâtiments

Si nous avons utilisé les bâtiments de la base de données OSM pour les calculs de part de surface occupée par d'autres objets de la base de données OSM, le critère évaluant la hauteur des bâtiments est construit à partir des données diffusées par l'APUR. Plusieurs variables portent des informations relatives à la hauteur de chaque bâtiment. Nous retenons la hauteur médiane calculée pour chaque bâtiment. Il s'agit d'une mesure qui exclut « les valeurs aberrantes » parfois obtenues à l'utilisation de modèles numériques de terrain. Selon notre critère, plus la moyenne des hauteurs médianes des bâtiments par cellule est faible plus la cellule obtient un nombre de points importants. Ce critère correspond à la volonté de proposer des espaces ouverts où se distinguent des perspectives.

- Accès à une terrasse

Dans la continuité des espaces de promenade, également propice aux haltes, nous ajoutons l'offre de lieux de restauration en terrasse. Ils sont à la fois, lieux de repos et points d'observation. Le jeu de données diffusé par la Mairie de Paris renseigne sur ce type d'objets, sous la description « terrasse ouverte ». Nous rapportons la surface des terrasses par cellule à la surface de la cellule à laquelle nous soustrayons la surface des bâtiments dessinés par l'APUR. D'après l'image C de la figure 58 (page 277), les terrasses sont surreprésentées au centre de la ville. On remarque également une forme assez proche de la répartition de la densité des voies piétonnes. Le nord-ouest parisien se distingue cependant, alors que les voies piétonnes y sont sous-représentées.

- Proximité d'arrêts de bus

La promenade urbaine à Paris peut se réaliser à pieds ou en bus. Notre dernier critère est donc la densité d'arrêts de bus par cellule. Les voies piétonnes ne peuvent pas être empruntées même par un moyen de transport en commun. Le calcul de densité prend donc en compte, par cellule, le nombre d'arrêt de bus en fonction des mètres de routes praticables. L'image I de la figure 60 met en évidence que l'accessibilité au réseau de bus est surreprésentée aux alentours de la gare Saint-Lazare (à l'ouest de la rive droite), de la gare de l'Est (au nord de la rive droite) et de la gare d'Austerlitz (au nord de la rive gauche). On remarque également une accessibilité supérieure à la moyenne dans le Bois de Vincennes.

Les points que nous affectons à chaque cellule sont relatifs à la position statistique de chacun des six critères par cellule. Nous cherchons à valoriser les cellules qui connaissent une surreprésentation de la variable mesurée. Ainsi, la première classe de discrétisation de chaque variable s'étend du minimum à la médiane et n'apporte aucun point. De ce fait, pour chaque variable, la moitié des cellules ne rapportent aucun point. Par exemple, une cellule présentant moins de 5% d'espace de promenade, n'obtient pas de point pour ce critère (fig. 57). Passé la médiane jusqu'au 72,5ème centile, non compris, la cellule obtient un point, du 72,5ème centile au 95ème centile, non compris, la cellule obtient deux points et une cellule présentant une valeur supérieure au 95ème centile compris obtient trois points (fig. 57). Cette discrétisation est inversée pour la variable de la hauteur moyenne des bâtiments. Passé la médiane jusqu'au 22,5ème centile, non compris, la cellule obtient un point, du 22,5ème centile au 5ème centile,

non compris, la cellule obtient deux points (fig. 57). Les trois critères de présence apportent systématiquement chacun un point.

Distributions des neufs variables pour l'ensemble des cellules de l'emprise

Centiles / Critères	Médiane	72,5ème	95ème	Max
Promenade	5%	11%	100%	-
Patrimoine	0%	3%	29%	98%
Terrasse	0%	1%	5%	24%
Seine et canaux	Présence			
Accessibilité au réseau de bus	1‰	2‰	4‰	21‰
Street Art	Présence			
Accès Wifi	Présence			
Voies piétonnes	1%	3%	13%	46%
Points croissants	-	1	2	3
Centiles / Critères	**Médiane**	**22,5ème**	**5ème**	**Min**
Hateur moyenne bâti	15m	12m	0m	0m
Points décroissants	1	2	-	-

Figure 57 Résumé statistique des neuf variables et affectation des points (Hayat, 2019)

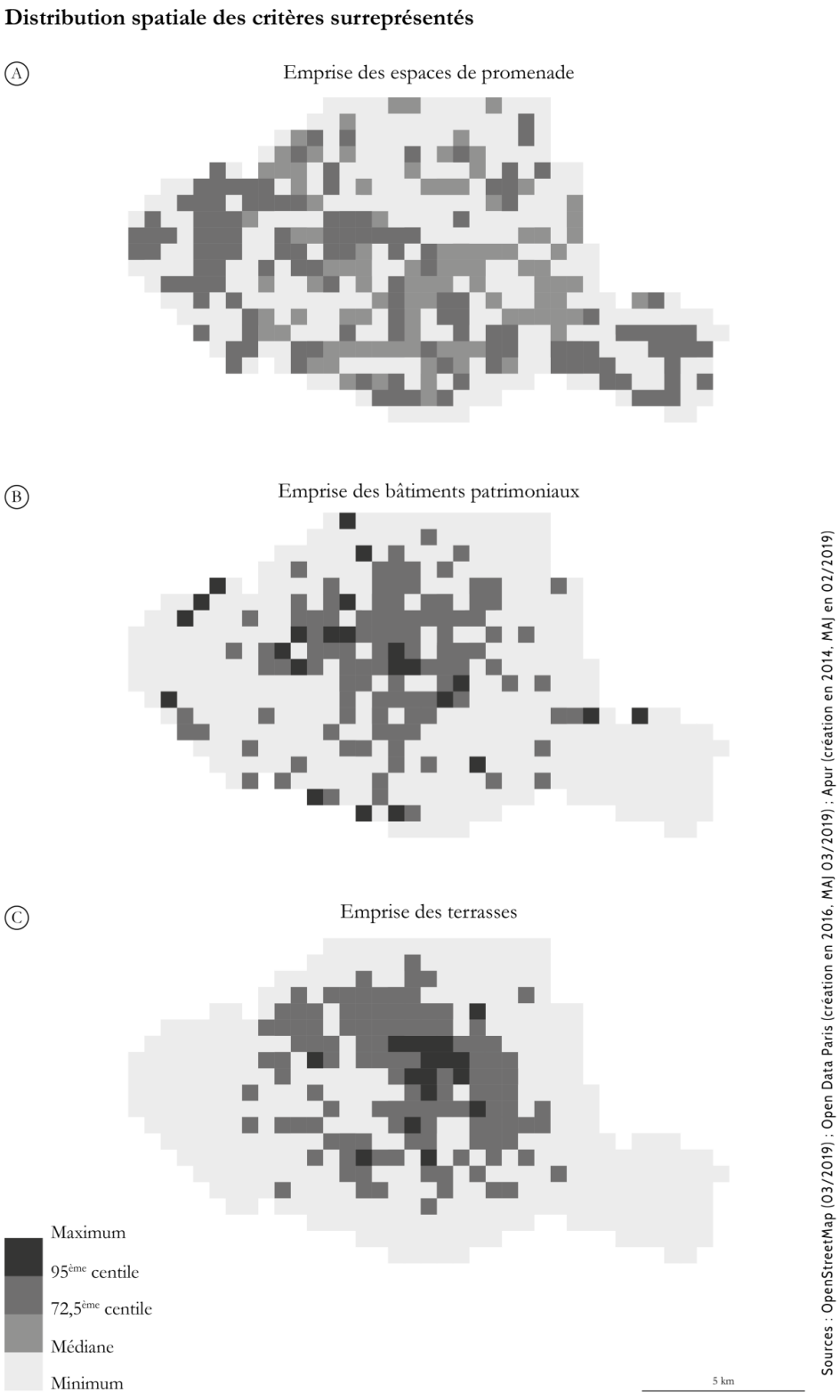

Figure 58 Résumés statistiques des variables emprise des espaces de promenade, emprise des bâtiments patrimoniaux et emprise des terrasses (Hayat, 2019)

Figure 59 Cellules qui comprennent au moins un service de Wifi, une photographie de Street Art ou une surface en eau (Hayat, 2019)

Distribution spatiale des critères surreprésentés

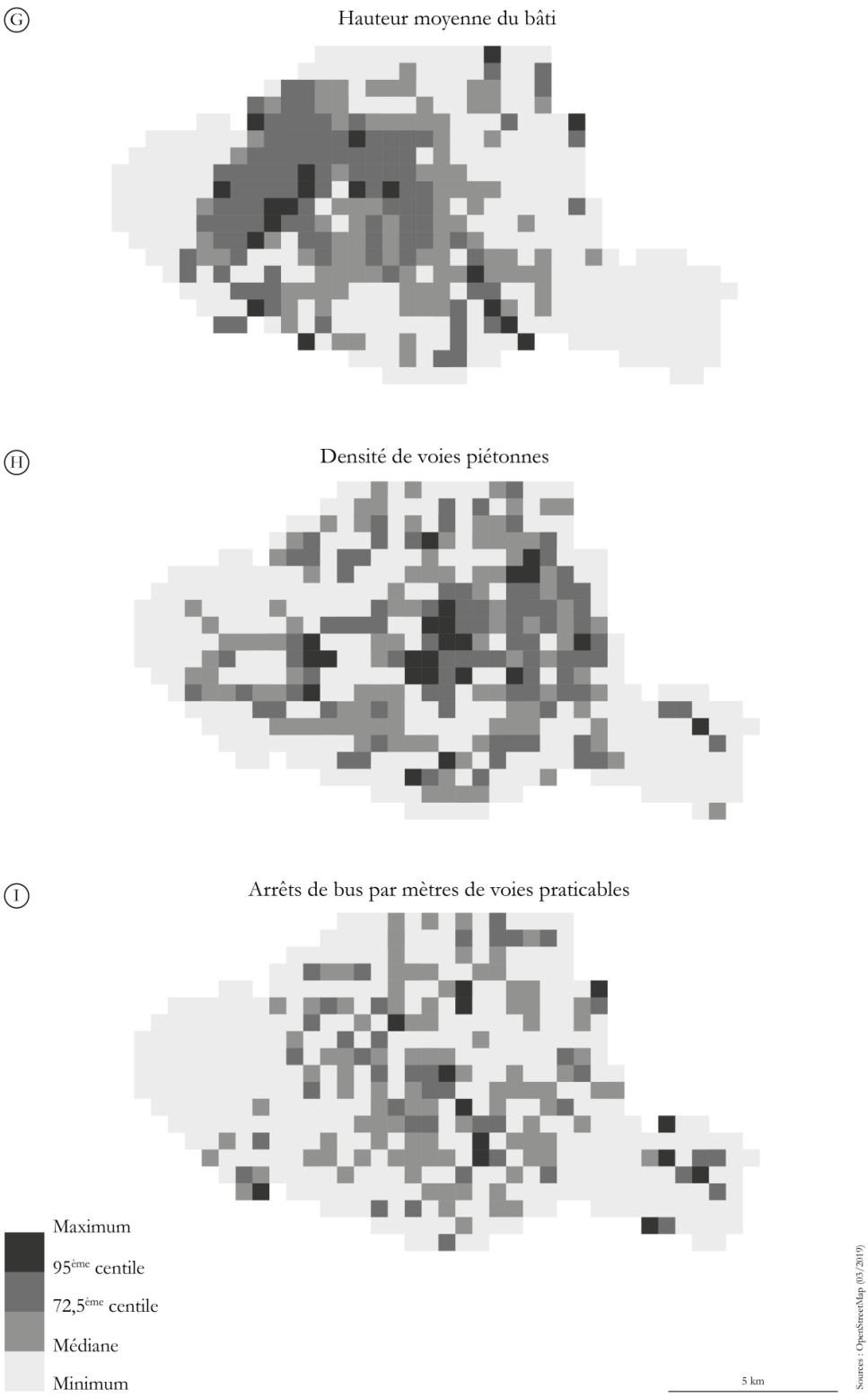

Figure 60 Résumés statistiques des variables hauteur moyenne du bâti, densité de voies piétonnes et densité des arrêts de bus (Hayat, 2019)

Nous résumons graphiquement une répartition possible de ces neuf critères en figure 61. Nous nous situons dans un cas théorique d'une cellule qui comprendrait l'ensemble des neuf critères. Compte tenu de la surface des cellules cette situation n'est que théorique. Elle nous permet cependant d'exposer le procédé d'affectation d'un rang à cette cellule en fonction des résultats réels que nous avons obtenus (figure 62).

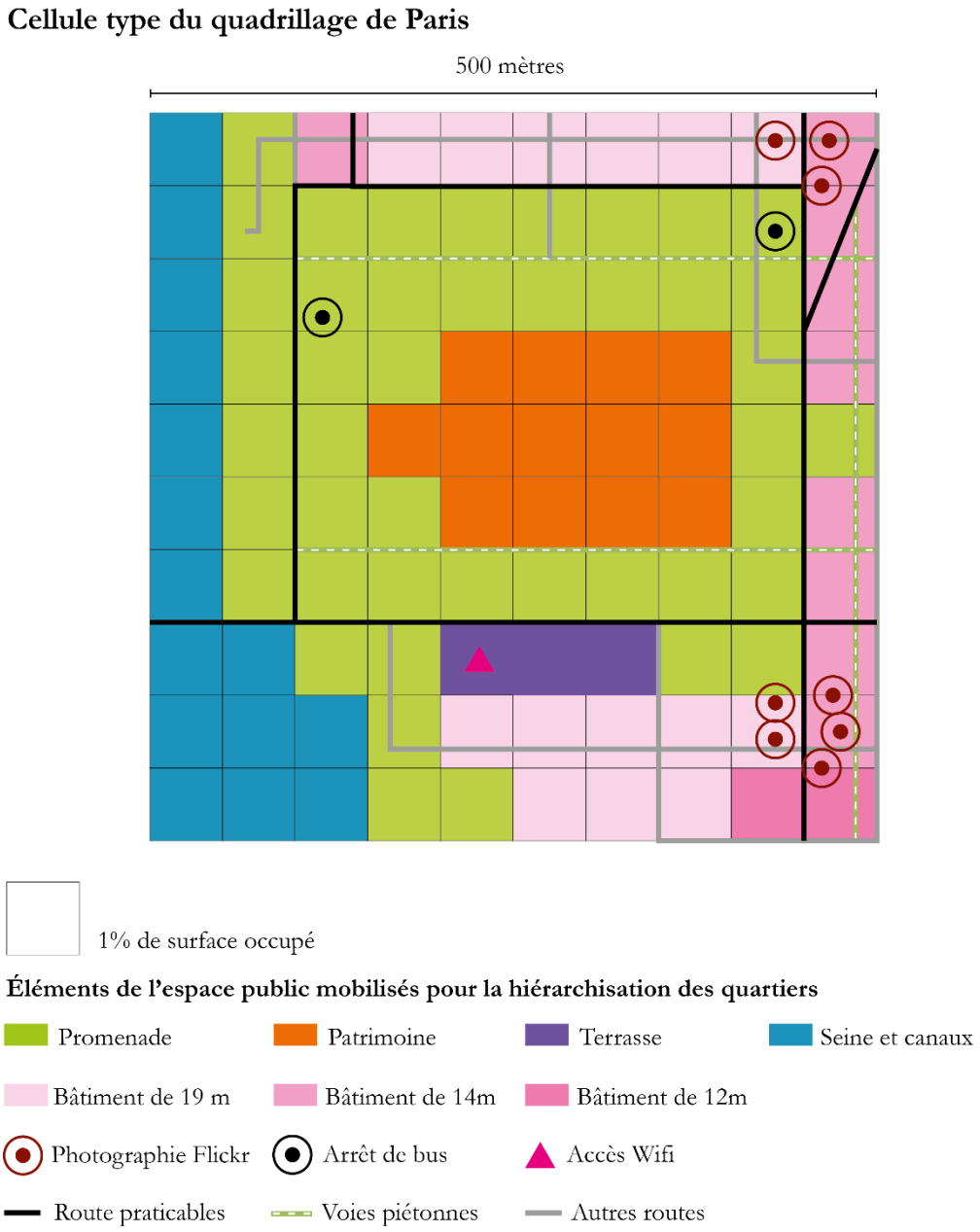

Figure 61 Les critères retenus en une possible répartition dans une cellule de la grille (Hayat, 2019)

Cellule type du quadrillage de Paris

Centiles / Critères	Valeurs	Bornes de classifcation	Points
Promenade	46%	11%<valeur≤100%	2
Patrimoine	13%	3%<valeur≤29%	2
Terrasse	3%	1%<valeur≤5	2
Seine et canaux	Présence		1
Hateur moyenne bâti	16,64m	16m<valeur	0
Accessibilité au réseau de bus	1‰	1‰<valeur≤2‰	1
Street Art	Présence		1
Accès Wifi	Présence		1
Voies piétonnes	24%	13%<valeur≤46%	3
Total			**13**

Figure 62 La cellule théorique obtient 16 points et se positionne donc parmi les cellules les plus recommandables (Hayat, 2019)

Au terme de l'affectation des points, le total est un indicateur de correspondance au quartier théorique de promenade que nous recherchons dans cet exemple. Nous cartographions cet indicateur dans la perspective de proposer une représentation grand public d'accompagnement au tourisme en ville. Nous discrétisons d'abord cette nouvelle variable selon les quartiles, dans la perspective de concevoir un ensemble de cartes comparables. En outre, notre propos est de mettre en évidence des zones attractives, ce sont les valeurs au-dessus de la médiane qui nous intéressent donc Nous ajoutons le seuil du 95ème centile afin d'insister sur les cellules à l'extrémité de la distribution. Nous obtenons alors une discrétisation en cinq classes (fig. 63).

Distributions statistiques des points cumulés

Centiles	25ème	50ème	72,5ème	95ème	Max
Points cumulés	3	5	7	10	15

Discrétisation de la variable des points cumulés pour sa cartographie

Bornes	valeur≤3	3<valeur≤5	5<valeur≤7	7<valeur≤10	valeur>10
Classes	1	2	3	4	5

Figure 63 Construction de la légende de la carte de synthèse des neuf critères (Hayat, 2019)

Le résultat obtenu (fig. 64), présente, au premier regard la géographie classique du Paris touristique. Les arrondissements centraux que sont le 1er, le 3ème, le 4ème, le 5ème et le quartier des Champs-Élysées concentrent le plus de cellules appartenant à la dernière classe. Nous choisissons d'ailleurs de représenter le quartier des Halles (à l'intersection de quatre cellules obtenant entre 9 à 15 points) et de Saint-Michel (centré entre deux cellules de 12 et 11 points) comme des exemples de plans des espaces les plus propices à la promenade urbaine (fig. 65) d'après notre chaîne de production. Nous notons cependant la mise en évidence des bords du canal Saint-Martin, des abords du parc des Buttes Chaumont, le nord du quartier de la Villette, le quartier des Ternes et les bords de Seine à l'est du quartier de la Salpêtrière, quartiers plus périphériques et davantage fréquentés par les parisiens que par les visiteurs de passage.

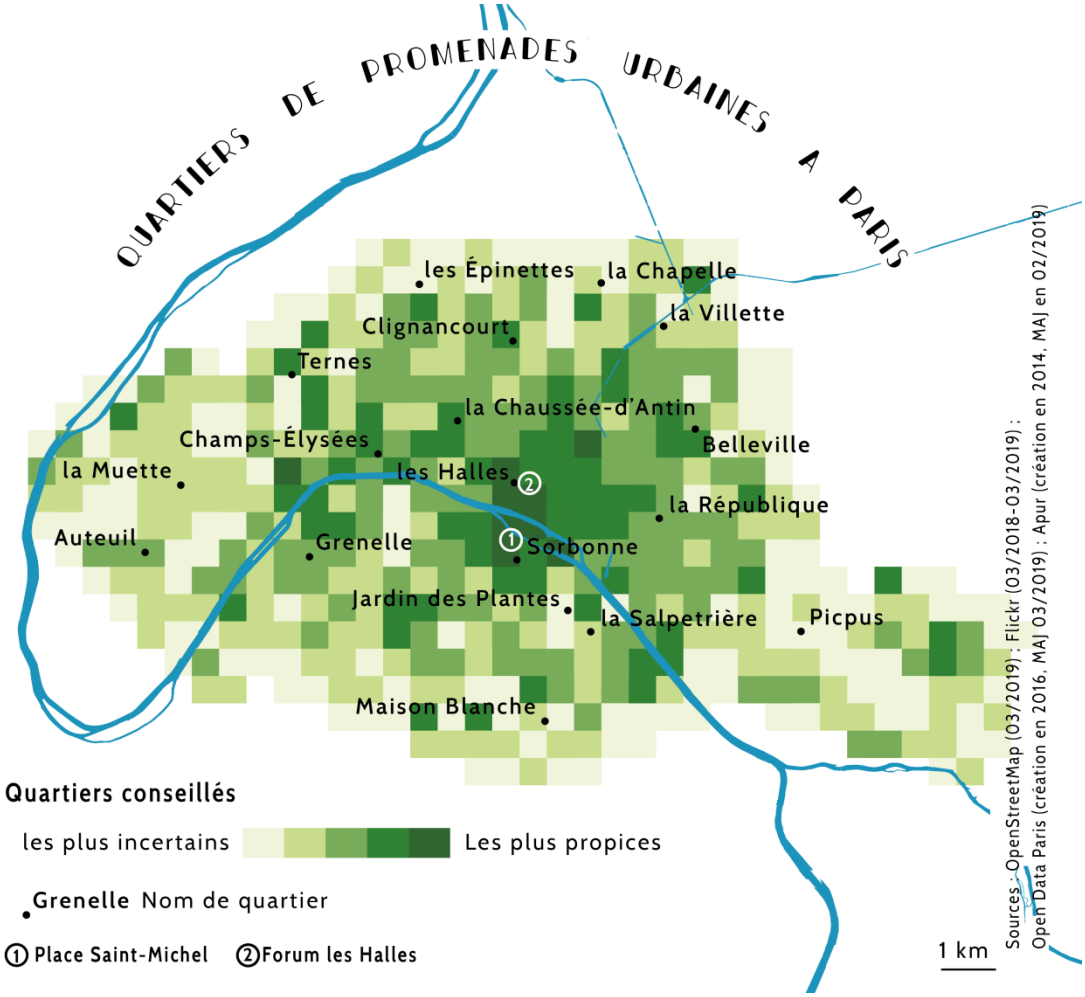

Figure 64 Cartographie du classement des cellules selon les neuf critères d'attractivité (Hayat, 2019)

Cet examen répond aux deux questions initiales. Le résultat, à l'état de prototype, est issu d'une chaîne de production conçue à partir de plusieurs sources de données ouvertes, compatibles, en partie institutionnelles et en partie relevant de la folksonomie. En outre, après avoir associé ces sources de données, nous sommes en mesure de proposer une représentation de la ville de Paris propice à la pratique du tourisme en bénéficiant des informations originales fournies par des communautés (telle que la présence d'œuvres de *Street Art*) ainsi que de la précision institutionnelle (telles que la hauteur du bâti et la surface des terrasses).

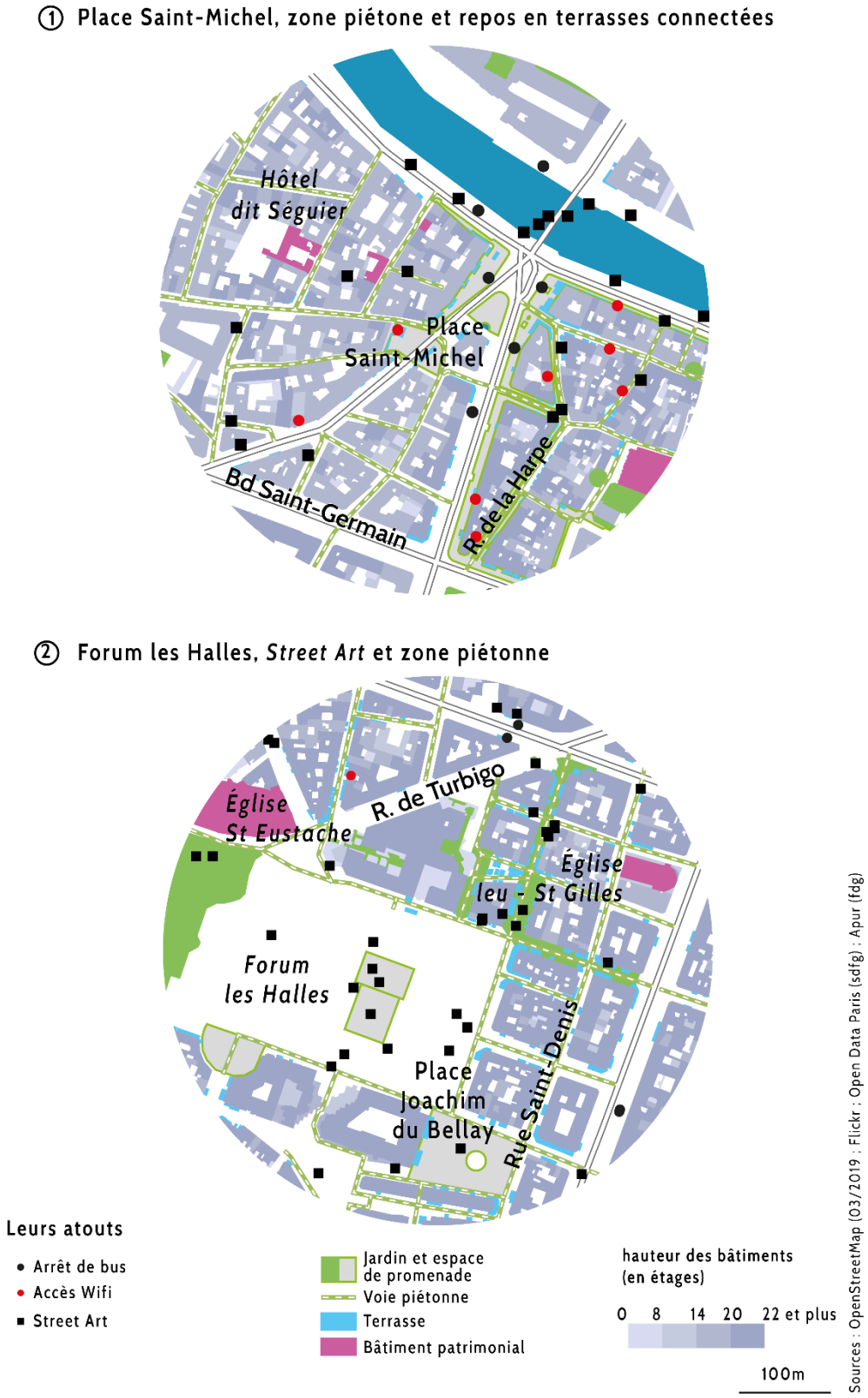

Figure 65 Exemple de plan rapproché de cellules qui obtiennent un rang élevé (Hayat, 2019)

Conclusion du chapitre 3

Mon intégration à l'entreprise Michelin dans le cadre du contrat CIFRE a été motivée par un intérêt prononcé pour les alternatives aux systèmes de production contemporains propriétaires d'exploitation d'informations géographiques pour la cartographie. En 2016, au sein du service de cartographie de Michelin, plusieurs chaînes de production se côtoyaient pour la production papier et chacune ne mobilisait quasiment uniquement que des logiciels et des sources de données propriétaires. Durant les trois années de travail, en une équipe réunissant des cartographes, des géomaticiens et des développeurs, il a été question de remodeler les chaînes de production afin de profiter pleinement des opportunités de production offertes par les principes du Libre appliqués aux ressources documentaires et au développement logiciel. Dans la première sous-partie de ce chapitre, nous avons expliqué les conséquences techniques, éditoriales, méthodologiques et communicationnelles d'une telle transformation. Nous avons retracé le processus de construction de chaînes de production nécessaire à l'exploitation des données OSM. Cette nouvelle chaîne n'est pas la reproduction parfaite des précédentes mais c'est le respect des mêmes attendus éditoriaux et géographiques qui a présidé à sa construction. Nous avons expliqué, du point de vue d'un producteur de cartes traditionnel, les avantages que représente le tournant du Libre. Les données utilisées font partie d'un ensemble de ressources ouvertes qui tendent à devenir interopérables. Pour les données comme pour les logiciels, le producteur peut devenir un acteur de ses développements. Pour une entreprise comme Michelin, le retour à la création d'information géographique est une opportunité éditoriale. Michelin peut ainsi bénéficier des contributions libres et potentiellement originales et apporter à la base de données ses propres informations. Cependant, le changement de modèle de production vers davantage de produits ouverts n'est pas dénué d'inconvénients. Dans le deuxième chapitre nous avons observé l'hétérogénéité des données et nous l'avons décrite comme une caractéristique inhérente au projet libre OSM. Il s'agit d'une des manifestations de la relative instabilité de projets libres. Dans le cas de simples outils tels que Imposm3 ou Utilery, l'abandon par son créateur et d'éventuels contributeurs est à envisager. Cela ne signifie pas la disparition de l'outil sur Internet, mais l'entreprise utilisatrice de tels outils se voit obligée d'en poursuivre le maintien si la chaîne de production en est dépendante et qu'il n'existe pas d'outil alternatif compatible. Cette notion d'alternative

est cependant fondamentale dans les principes du Libre puisque c'est en effet, en réaction à la fermeture de certains produits logiciels que ce mouvement s'est manifesté.

Dans la deuxième sous-partie du chapitre, nous avons exploré l'émergence de nouveaux acteurs du domaine de la cartographie et de l'exploitation d'information géographique qui se présentent en alternatives à Google Maps, Bing Maps et TomTom. Les acteurs dont nous étudions l'activité, sont, à la différence de Michelin, majoritairement nés suite à la diffusion des informations libres. Une dynamique commerciale intense compose l'environnement du projet OSM. Au terme de cette deuxième partie, nous rejoignons les conclusions de Leon et Quinn (2018) selon lesquelles cette dynamique profonde vient perturber l'équilibre fragile et très récent du projet OpenStreetMap. Nous explorons dans le chapitre quatre, les premières conséquences de ces activités économiques nouvelles.

Enfin, en troisième sous-partie, nous nous sommes placés dans le rôle d'un nouvel acteur de la cartographie touristique. Nous avons testé la faisabilité de l'utilisation de ressources libres pour la production de cartes à la fois originales et fiables. La compatibilité des licences encourage à la mise en relation de sources documentaires si distinctes qu'elles ne peuvent être présentes dans les ressources d'un même fournisseur. Si la réalisation est possible, l'interopérabilité des sources n'est pas garantie. Alors qu'on ne peut attendre d'une ressource comme Flickr qu'elle soit même modélisée, en tant que plateforme de partage de photographies simplement *taggés* sans discussion relative au bon usage des *tags*, une croissance des relations entre les sources institutionnelles d'une part et les sources institutionnelles OSM d'autre part est envisageable.

Le principe d'ouverture des logiciels et des ressources documentaires stimule la conception d'alternatives à des services commerciaux largement dominants. La position quasi hégémonique de la société Google s'explique par l'anticipation et le développement des besoins associés à la nouvelle technologie d'Internet. Le Libre et les entreprises qui construisent leur système économique sur les atouts et les contraintes de ce courant se diffusent également par ce réseau. Les relations de concurrence entre nouveaux acteurs nés avec Internet concernent désormais les acteurs traditionnels. Ces derniers s'intègrent progressivement à Internet non pas simplement pour publier des

produits finis mais également comme acteurs des relations sociales et techniques qui émergent sur Internet.

Nous avons exposé la place des entreprises au sein des communs qui composent l'écosystème d'OSM. Dans le chapitre suivant, nous interrogeons la cohérence du système tel qu'il est par rapport aux règles établies par les communautés d'OSM et du Libre. Ce chapitre permettra ainsi de clore l'analyse poursuivie dans cette thèse Dans ce dernier chapitre, nous exposons le rôle des innovations techniques, communicationnelles et géomatiques ainsi que le rôle des communautés en ligne sur la réalisation du projet OSM.

Chapitre 4 : Quel avenir pour l'Information géographique libre ?

L'objet de ce chapitre est de **proposer une analyse qui restitue des éléments explicatifs du dispositif socio-technique OpenStreetMap**, sans se limiter aux seuls éléments spécifiques de l'infrastructure d'OSM.

Dans le premier chapitre, nous avons considéré OSM comme un cadre spécifique de production de données géographiques : un cadre qui relève du Libre et s'intègre dans la grande catégorie de production par *crowdsourcing*. Dans le deuxième chapitre, nous avons rendu compte de l'intégration progressive d'OSM dans des publications institutionnelles, grâce à une reconnaissance de la fiabilité des données produites dans les contextes qui l'ont mobilisé. L'utilisabilité technique et thématique des données a été explorée dans le troisième chapitre. Ce chapitre nous a également permis d'observer la naissance d'entreprises spécialisées dans l'exploitation des données contributives et l'émergence de nouvelles concurrences entre les acteurs de l'information géographique. Nous avons pu faire la preuve de l'efficacité de l'écosystème technique interopérable d'OSM par la construction d'une cartographie touristique industrielle. Dans ce quatrième chapitre, nous abordons maintenant la question des relations entre le développement de la pratique du *crowdsourcing*, la diffusion de la pensée du Libre et l'exploitation commerciale des données OSM. Ces trois éléments du système OSM prennent une place croissante au sein du réseau Internet. Nous nous interrogeons sur les perspectives qui peuvent être envisagées pour une production cartographique et les usages des données OSM. Notre perspective est exploratoire, à la recherche d'indices qui montreraient la croissance des projets libres au cours des quinze dernières années. Ces indices nous permettront d'envisager des évolutions à venir, de considérer la dynamique initiée pour une information géographique non contrainte et toujours plus librement accessible.

Nous évoquons en première partie du chapitre, les développements simultanés de différentes techniques de partage d'informations et de leurs enregistrements. Les formes de sociabilité qui se sont constituées autour du projet OSM ne peuvent se comprendre que par rapport aux dispositifs de communication préexistants sur Internet. Enfin, nous terminons cette première sous-partie du chapitre en exposant les transformations rapides des relations sociales et professionnelles qui se tissent au sein

du projet OSM. Le *crowsourcing* est à la fois un processus de création bénévole en communautés et un processus de captation d'informations produites par les internautes, parfois rémunéré, par des acteurs économiques dominants sur le réseau Internet. Alors que le Libre a été une réaction à l'appropriation de connaissances numériques jusqu'alors mises en commun, il semble que le phénomène d'appropriation se reproduise dans les années 2010 et mène à des comportements concurrentiels au sein même des communautés d'internautes. De tels comportements semblent pourtant contrevenir au sens du projet initial. Se pose ainsi la question de la faisabilité du concept de communs numériques par rapport à celui de communs posés en 1990 par Elinor Ostrom. Les évolutions dans les interactions au sein d'OSM se trouvent également parmi les entreprises qui pour certaines prennent un rôle dirigiste sur la gestion du commun numérique.

Si aujourd'hui la pratique de la contribution à la construction de base de données géographique peut être un travail rémunéré mais aussi un travail bénévole, vers quel prochain équilibre l'écosystème OSM se dirige-t-il ? Alors que les acteurs industriels sont des contributeurs, aussi bien à l'infrastructure qu'à la base de données, leur influence dirigiste perturbe-t-elle le processus de consensus établit par les contributeurs individuels ? Nous pouvons nous demander si la distinction entre contributeurs individuels influents et acteurs industriels est toujours d'actualité. En effet, les acteurs industriels du domaine communiquent en ligne par la voix d'employés identifiés comme porte-parole de leur entreprise et habiles techniciens qui prennent le rôle de contributeur individuel.

4.1. Internet, un contexte spécifique pour de nouveaux cadre de production des données

Dans cette sous partie nous cherchons à mettre en évidence la relation de deux domaines d'innovation : l'invention de nouvelles techniques et leur perfectionnement par le moyen de nouveaux systèmes de communication. Nous testons ainsi l'hypothèse selon laquelle les nouveaux moyens de communication ont été les berceaux d'autres innovations dans des domaines associées aux nouvelles technologies comme l'information géographique. Celle-ci est aujourd'hui produite, gérer et éditer au moyen de techniques numériques qui ont commencé à voir le jour au début des années 1960 (fig. 66). Ainsi, si la conception des premiers systèmes d'informations géographiques précède la mise en réseau du monde par le moyen d'Internet, les années 1960 sont également les années du développement de projets comme ARPANET aux États-Unis et Cyclades en France, deux projets d'échange de documents par la mise en réseau d'ordinateurs. L'apparition des domaines des « protocoles et procédés de communication » et des « technologies des SIG » se sont produits simultanément, du milieu des années 1960 au milieu des années 1970 (fig. 66). À ces deux domaines, s'ajoutent à partir des années 1990 et l'invention du World Wide Web, la création de moyens permettant la mise en partage de « services en ligne d'échanges » (tels que le protocole http et le langage PHP). À la fin des années 1990, un réseau de communication dynamique et mondial était créé. Au milieu des années 2000, l'information géographique prenait alors une place centrale sur ce réseau suite à l'ouverture du réseau GPS et à l'usage massif des images (donc des cartes) comme média d'informations. Ces nouveaux domaines de création technique sont portés par des projets de recherches et sont également rapidement des objets d'études pour les scientifiques. Il faut signaler que tous ces bouleversements sont associés à l'évolution des moyens matériels de l'informatique qui se sont accrus de manière exponentielle depuis l'invention des premiers ordinateurs, même si nous n'y ferons pas spécifiquement mention ici.

La figure 66 est une synthèse d'éléments que nous avons sélectionnés comme représentatifs des techniques et des événements sociaux et économiques qui constituent les briques de l'Internet contributif. Elle n'a pas vocation à être exhaustive car nous ne retenons que les éléments qui nous semblent les plus fortement explicatifs de l'émergence du projet OSM et de son intégration dans le système économique de

l'information géographique. Mais elle nous permet de synthétiser notre propos et d'exposer la simultanéité de certaines inventions et d'entreprises majeures et la proximité des concepts scientifiques et leurs réalités sociétales. Nous commentons donc ce tableau en deux parties essentielles de la compréhension de l'évolution du système OSM, de son mode de production à sa construction sociotechnique au sein d'Internet.

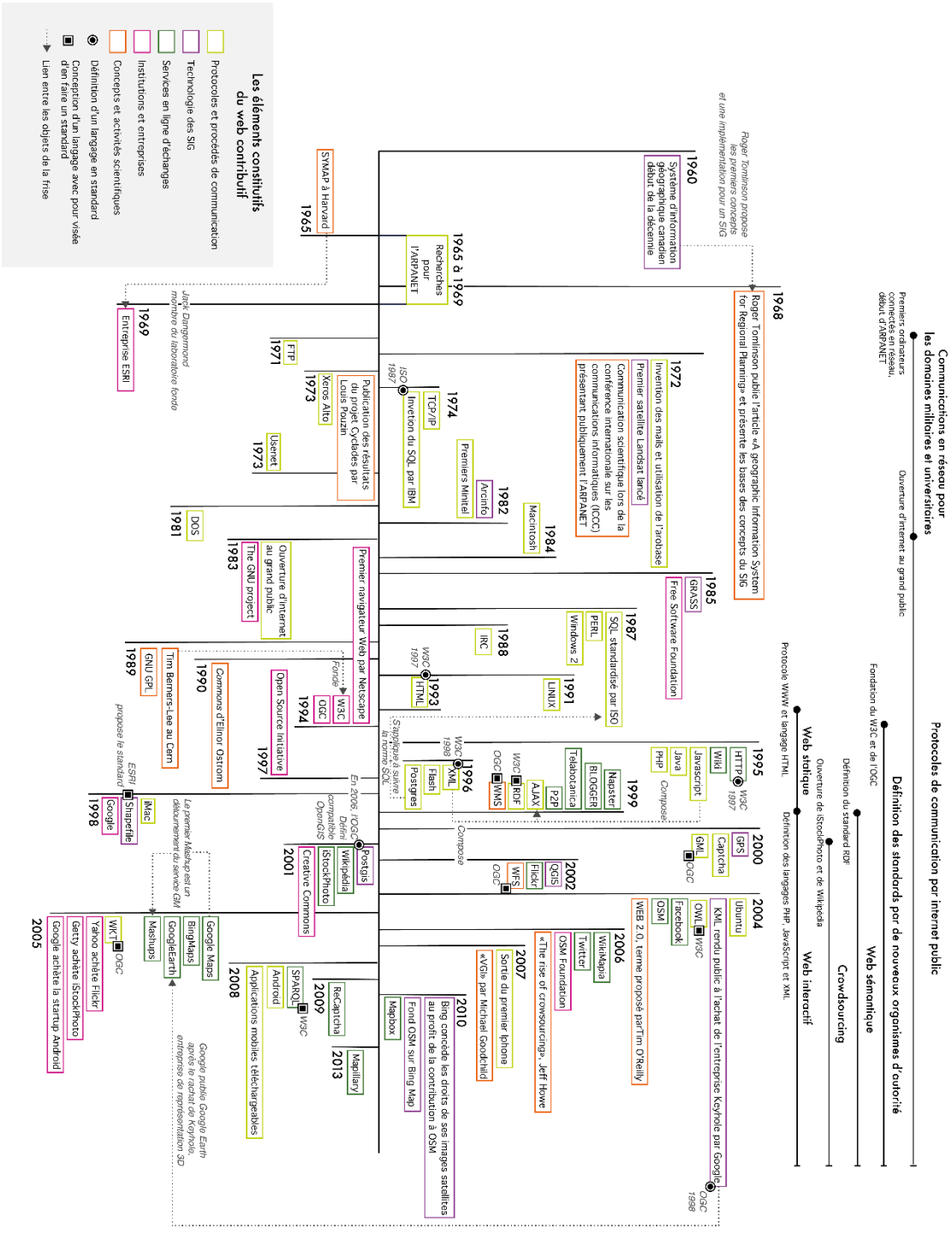

Figure 66 Frise chronologique d'éléments représentatifs facteurs de l'émergence du projet OSM (Hayat, 2019)

4.1.1. Des modes de production simultanés

Nous tentons d'exposer ici les éléments constitutifs du mode de production des connaissances libres, depuis les premiers développements technologiques et conceptuels de cette période, jusqu'à la fin des années 2010. Des années 1960 aux années 1990, le réseau Internet est construit essentiellement aux États-Unis par une communauté scientifique composée de militaires et d'universitaires. Des années 1960 au début des années 1980, le réseau Internet est destiné au partage de connaissances entre chercheurs.

Dans l'article « *A Brief History of the Internet* », co-écrit par neuf chercheurs américains fondateurs des technologies d'ARPANET et d'Internet, les auteurs décrivent Internet comme « *a widespread information infrastructure* » (Leiner et *al.*, 2009). Ils développent leur théorie sur les origines et l'histoire du développement d'Internet des années 1960 aux années 1990. Cette publication permet de comprendre comment cette période historique est fondamentale dans le développement de la contribution et du développement de l'information géographique libre. Elle commence avec la publication de l'article « *Information Flow in Large Communication Nets* », en 1961, par Leonard Kleinrock. Celui-ci conçoit la technique d'échanges de documents par « la commutation de paquets », ce qui deviendra le vrai socle technique d'Internet. La période décrite s'achève avec la définition, du réseau Internet par le Federal Networking Council (1995) : « *Internet » refers to the global information system that -- (i) is logically linked together by a globally unique address space based on the Internet Protocol (IP) or its subsequent extensions/follow-ons; (ii) is able to support communications using the Transmission Control Protocol/Internet Protocol (TCP/IP) suite or its subsequent extensions/follow-ons, and/or other IP- compatible protocols; and (iii) provides, uses or makes accessible, either publicly or privately, high level services layered on the communications and related infrastructure described herein* »[226]. Corroborant cette définition, les neuf auteurs-inventeurs présentent la principale et première motivation du développement d'ARPANET et d'Internet comme étant le

[226] « Internet »fait référence au système d'information mondial qui: (i) est lié de manière logique par un moyen d'adressage unique basé sur le protocole Internet (IP); (ii) est capable de prendre en charge les communications à l'aide de l'ensemble des protocoles TCP / IP (Transmission Control Protocol / Internet Protocol) ou d'autres protocoles compatibles IP; et (iii) fournit, utilise ou rend accessible, de manière publique ou privée, des services de haut niveau associés aux infrastructures de communication. (Traduction par l'auteure.)

partage de ressources. Ces ressources prennent deux formes : le partage de matériels coûteux pour le calcul à distance et le partage d'informations. Les informations partagées dans ce cadre consistent en une documentation entre membres de la « communauté scientifique » composée des chercheurs de différentes universités concernées par le projet ARPANET (MIT, UCLA, Standford). Les auteurs considèrent d'ailleurs que « *a key to the rapid growth of the Internet has been the free and open access to the basic documents, especially the specifications of the protocols* »[227]. Cependant, considérant que l'échange de connaissances par le moyen traditionnel de publications est trop lent, les chercheurs développent un canal de communication, d'abord sous format papier et postal nommé *Request for Comments* (RFC ou demande de commentaires). Il s'agit de mémos informels diffusés rapidement aux membres de la communauté. Le protocole d'échange de document numérique FTP[228] (1971) remplace ensuite le courrier postal pour ces échanges de connaissances. Le développement des moyens de communication, au service du travail scientifique, se poursuit et l'e-mail est inventé (1972, fig. 66). Les auteurs de « *A Brief History of the Internet* » considèrent que le *mail (*ou courrier électronique) a fourni un nouveau modèle de communication entre les personnes et a modifié la nature même de la collaboration des inventeurs du réseau Internet (Leiner et *al.*, 2009). Ils décrivent le mail comme un mécanisme de coordination simple qui a soutenu le développement rapide d'internet puis, qui est devenu un outil de collaboration pour « une grande partie de la société ».

L'histoire d'Internet présente ainsi quatre aspects bien différents : une révolution technologique (la commutation de paquets), qui s'est développée depuis en une infrastructure à plusieurs dimensions (i) ; un système de gestion pour la coordination qui, à partir des années 1990, tend vers une coordination mondiale (ii). Ces deux aspects ont un impact fort sur la société et l'exemple du mail est probant dans cette affirmation de la porosité de la société à Internet (iii)). Enfin, Internet a pris une ampleur mondiale, en partie du fait de sa facette commerciale (iv). Les entreprises déjà en place dans les domaines de l'informatique se sont rapidement concentrées sur les technologies d'Internet et de nouvelles entreprises ont été fondées, avec, à leur tête, d'anciens chercheurs ou étudiants des universités où se sont décidés les premiers protocoles de

[227] L'accès libre et ouvert aux documents de base, en particulier aux spécifications des protocoles, a été l'un des facteurs clés de la croissance rapide d'Internet. (Traduction par l'auteure.)
[228] *File Transfer Protocol* (protocole de transfert de fichiers).

communication. Ainsi, le mail reste une méthode basée sur un protocole d'échange de fichiers qui a été conçue pour répondre à un besoin lié à l'ampleur du réseau et à l'ambition d'un développement très rapide de la technologie du partage de fichiers en réseau. Ce système de communication est aujourd'hui un moyen quotidien d'échanges entre particuliers et entre entreprises, à l'échelle mondiale. C'est cette plasticité de la technologie d'Internet qui est observée par les auteurs, inventeurs du réseau Internet, lorsqu'ils considèrent que son expansion rapide actuelle est alimentée par la réalisation de sa capacité à promouvoir le partage d'informations. Ils rappellent également le concept-clé d'Internet qui consiste en une architecture générale à laquelle différentes applications peuvent être connectées. Dans la mesure où Internet n'a pas été conçu pour une unique application, ses concepteurs ont laissé la place à l'invention du concept de World Wide Web (Web).

Nous intercalons, à cette description de l'évolution d'Internet, que nous avons présenté comme plutôt américaine, l'histoire du programme Télétel et de son terminal, le Minitel. L'étude de ce réseau nous apporte des éléments pour comprendre le développement du contributif dans le cadre du Web.

4.1.1.1. Le minitel

Le Minitel est un outil *télématique* (Nora-Minc, 1977) conçu par le Centre national d'études des télécommunications et le Centre commun d'études de télévision et télécommunications. Il est installé en France, dans les foyers et les entreprises, à partir 1982 et jusqu'aux années 2000. Ce média, largement implanté en France, est supplanté par le Web au tournant des années 2000 et a finalement été abandonné le 30 juin 2012 par l'État et son opérateur, les PTT (devenu France Télécom à la fin des années 1980 et Orange en 2013).

Bien qu'aujourd'hui obsolète, le Minitel était, dans la France des années 1980, une innovation technologique sans pareille[229]. L'appareil est construit comme un objet complet comportant un terminal, un écran et un clavier. L'ergonomie est tournée vers

[229] L'invention et la diffusion du Minitel répondait à un projet politique national, dans un contexte de forts investissements engagés par le pouvoir central. En 1978, le Président Valéry Giscard d'Estaing a décidé de renoncer au projet de recherches Cyclades développé par plusieurs centres de recherches et dirigé par Louis Pouzin. L'équipe construisait un système de mise en réseau d'ordinateurs selon le modèle conçu sous le nom de « commutation de paquets ». À l'« équivalent français de l'américain ARPANET dans le domaine informatique » (Cariou et Gaulon-Brain, 2012) Valéry Giscard d'Estaing a préféré la simplicité du Minitel. La France s'est alors tournée vers un réseau centralisé.

la simplicité d'utilisation pour le plus grand nombre. Il n'est donc pas nécessaire de posséder un ordinateur personnel pour accéder à un réseau de communication. En France, très rares sont les foyers qui disposent d'un ordinateur personnel avant les années 1990. D'ailleurs, ils ne sont réellement produits, notamment aux États-Unis, à destination des particuliers, experts de l'informatique, que depuis le milieu des années 1970. À l'époque où le rapport Nora-Minc (1977)[230] expose la tendance globale à l'informatisation, la France connaît donc un faible taux de pénétration des ordinateurs : moins de 5 % des foyers sont équipés au début des années 1980 (Cariou, Gaulon-Brain, 2012)). Le gouvernement et les directions responsables des télécommunications dirigent les investissements vers un système complet afin de rapidement doter le pays d'un système de télécommunication. L'installation de ce réseau mobilise le réseau de téléphonie et bénéficie d'une large campagne d'installation nationale (Cariou, Gaulon-Brain, 2012).

L'outil doit d'abord permettre à tous les foyers d'accéder à l'annuaire téléphonique. Dans le contexte d'une large extension du réseau téléphonique, celui-ci est difficile à maintenir à jour en version papier : la télématique est perçue comme « la solution ». Il est conçu initialement une infrastructure de diffusion d'informations contrôlées par les institutions qui en sont responsables et non comme un simple système de communication. C'est pourquoi, dans le contexte du développement d'un service public, le terminal Minitel est gratuitement proposé aux foyers raccordés au réseau de téléphonie.

Avant la diffusion généralisée du service, une expérience locale visant à tester l'accessibilité technique du Minitel a été organisée. Cette expérience a transformé le concept initial du Minitel en détournant sa fonction. Les utilisateurs ont en effet inventé un service de messagerie à l'insu des responsables. Dans le cadre d'une l'expérience locale, menée à Strasbourg par le journal *Les dernières nouvelles d'Alsace* (DNA) et en partenariat avec la Direction Régionale des Telecom (DRT), l'expérience GRETEL, un dispositif est mis en place afin de mettre en relation usagers et concepteurs. Le service informatique des DNA décide de créer, dans le but d'accompagner les

[230] Publié en 1977, le rapport répond à une commande du Président de la République, Valéry Giscard d'Estaing. Celui-ci s'adresse à l'inspecteur général des finances Simon Nora. Nora s'associe à Alain Minc afin de proposer une « réflexion sur les moyens de conduire l'informatisation de la société » française. Les auteurs proposent le terme de télématique qu'ils définissent comme « l'imbrication croissante des ordinateurs et des télécommunication ». Les auteurs interrogent les conséquences de l'apparition de la télématique sur la société et soulignent les enjeux économiques de ce processus. Ils incitent à une prise de conscience et à une démocratisation de la télématique.

utilisateurs dans la découverte de l'outil, « une messagerie interne qui ne fonctionne que dans un sens, du centre serveur vers les usagers, elle nécessite une présence continuelle pour pouvoir répondre aux usagers » (Bruhat, 1984). Cette invention est rapidement détournée, voire « piratée », selon le point de vue adopté et une messagerie autonome des autorités (qu'on appellera plus tard la messagerie Gretel) est créée. Dans une étude réalisée par le SPES/DGT, intitulée « Genèse. GRETEL : la messagerie interactive - Histoire d'un piratage », Thierry Bruhat (1984) réunit les points de vue de la DRT, du service informatique des DNA et celui des utilisateurs à propos de ce piratage. Selon Thierry Bruhat, la DRT considère que « Gretel est née par hasard, ce n'était pas prévu... ». Le rapport précise qu'« il a été reproché [à la DRT ou aux DNA] de ne pas avoir contrôlé les opérations [...] les défoulements que la messagerie a entraîné » sans que l'on connaisse précisément l'origine des reproches. Nous supposons que ces reproches proviennent partiellement des usagers qui ont pu expérimenter un sentiment d' « intrusion de leur espace privé » (Bruhat, 1984) et reprochent un certain amateurisme du service informatique des *Dernières nouvelles d'Alsace* : « les DNA avaient fait un service de messagerie Internet et ils se sont fait pirater par des usagers » (Bruhat, 1984). Néanmoins ce que Thierry Bruhat décrit comme un « embryon d'interactivité » a finalement mené aux services de discussions en réseau qui ont fait la fortune de Xavier Niel, par exemple.

Au-delà de l'émergence de nouvelles entreprises, dans ce contexte de communication, France Télécom a largement profité de telles pratiques commerciales. En effet, si le Minitel était gratuit, ce n'était pas le cas de la fréquentation du réseau, tarifée en fonction du temps de fréquentation et dont la charge était reportée sur la facture de téléphone du particulier. Il rappelle que l'autre expérience locale, réalisée, en parallèle à Vélizy, a également nécessité un système d'échange entre l'opérateur et les usagers mais qu'il a pris la forme de courriers postaux. Nous retrouvons là une analogie avec l'invention des mails par les ingénieurs de l'ARPANET.

Au contraire, les PTT ont conçu un système fermé et homogène, à l'opposé du système d'Internet[231]. Si le réseau, connecté au Minitel, permettait un contrôle fort sur les

[231] Le Minitel est d'abord conçu comme un média pour les autorités de diffusion d'informations traditionnellement diffusées sous format papier. Cependant, le réseau s'est rapidement transformé, suite à un détournement technique par les usagers. Il a donc été conçu un réseau pour l'usage spécifique de la diffusion et non pour l'échange d'informations alors que celui-ci s'est rapidement transformé en ce sens. Cependant, le réseau n'était accessible que par le moyen du Minitel. Ce Minitel ne permettait pas le stockage ou le traitement de données et il était impossible d'y connecter des systèmes périphériques (physiques ou virtuels).

services, tenus de se déclarer aux PTT ou à France Télécom, et sur les échanges, il n'était pas adapté à l'affluence nouvelle des ordinateurs personnels entre la fin des années 1990 et les années 2000. Le Minitel n'est donc pas l'espace virtuel où la création d'informations contributives a pu voir le jour.

Nous continuons notre investigation afin de comprendre le développement du mode de production contributif et volontaire. Dans ce but, nous ouvrons le champ des recherches à chacun des éléments techniques et conceptuels qui ont pu participer à son émergence et à sa croissance. Comme nous l'avons vu au travers des exemples du mail et de la messagerie Gretel, les réseaux informatiques sont des moyens de communication qui encouragent l'échange continu et le partage dans une perspective de développement rapide des moyens de communication eux-mêmes. Nous procédons à un inventaire chronologique et sélectif des technologies de communication qui y participent. Nous identifions ces éléments comme associés au développement des communications entre deux personnes puis entre groupes et entre communautés.

4.1.1.2. Les protocoles techniques et sociaux de communications

Entre 1971 et 1988, sont inventés le FTP (1971), le mail (1972), la plateforme Usenet (1979) et IRC (1988). Ces quatre procédés de communication, différents d'un point de vue technique, sont comparables par le rôle qu'ils ont joué dans la constitution de communautés en ligne.

Le ***File Transfer Protocol*** **(FTP)** est un protocole de communication permettant le partage de fichiers entre ordinateurs connectés par le réseau Internet. Le partage consiste en une mise à disposition de documents depuis un ordinateur *client* sur un *serveur* accessible depuis Internet. Les personnes autorisées à se connecter à un serveur FTP ont le droit d'y enregistrer des documents, de copier les documents qui s'y trouvent ou de les supprimer (en fonction de leurs droits d'édition). Ce procédé de transfert a ouvert la possibilité de partage quasi instantané de documents, offrant également la possibilité à plusieurs utilisateurs, potentiellement distants de plusieurs milliers de

kilomètres, d'accéder aux mêmes documents simultanément. La *request for comments* 959 (RFC 959) d'octobre 1985 décrit les objectifs du protocole FTP[232] :

- promouvoir le partage de fichiers (programmes informatiques ou données) ;
- encourager l'utilisation d'ordinateurs distants ;
- protéger un utilisateur des variations des systèmes de stockage de fichiers en garantissant un système de stockage pérenne ;
- garantir l'efficacité du transfert des données de manière à ne pas altérer les données copiées.

Nous avons décrit le rôle des messageries instantanées et archivées dans le développement de ces mêmes technologies de communication et dans la construction de communautés partageant des intérêts communs. L'invention du **réseau Usenet** s'inscrit dans la suite logique de ces innovations. Ses inventeurs exploitent le concept du nouveau moyen de communication qu'est l'e-mail et créent un réseau distribué de groupes de nouvelles (*newsgroup*), organisés en fils de discussions. Le réseau mobilise les concepts de la technologie d'ARPANET et met en relation des universités entre elles. Camille Paloque-Berges (2012) décrit Usenet comme « l'ARPANET du pauvre », un type du réseau construit « pour soi-même ». Initié en 1979, par des étudiants de l'Université de Caroline du Nord, ce réseau, Usenet est une alternative à l'utilisation du réseau ARPANET à destinations d'universitaires nécessitant des moyens de télécommunications numériques mais non intégrés au réseau de scientifique des laboratoires de l'Arpa (Paloque-Berges, 2012). Usenet devient ensuite une plateforme de partage d'informations structurées en thématiques mais dont l'organisation sociale est *a priori* horizontale.

Elizabeth Reid (1991, 1996) et Michel Marcoccia (2001) présentent deux exemples de constitution de communauté autour de ces plateformes de discussion en ligne. Michel Marcoccia (2001) s'intéresse plus particulièrement à la plateforme Usenet et à l'émergence d'un « animateur informel » dans des forums non modérés. Elizabeth Reid, étudie quant à elle, dès 1991, **l'espace de discussion instantanée (chat) IRC** et expose la construction de nouvelles hiérarchies sociales au sein des communautés qui se sont constituées. La dernière technologie d'internet de notre liste qui nous semble avoir encouragé la constitution de communautés numériques est donc le chat IRC (Internet Relay Chat).

[232] https://www.ietf.org/rfc/rfc959.txt, [consulté en février 2019]

La plateforme de forums **Usenet** est un nouvel espace d'échanges. Les membres s'identifient par des pseudonymes et les distances géographiques ou sociales s'effacent. Les participants aux forums construisent des « documents numériques dynamiques, collectifs et interactifs » sous la forme des forums (Marcoccia, 2001). Chaque forum est construit par une communauté, réunie autour d'une thématique à laquelle chacun est sensible. Il en est de même pour le **chat IRC**, davantage tourné vers l'immédiateté des échanges mais dont les discussions sont archivées et peuvent servir de documentations. En effet, Elizabeth Reid (1996) identifie la constitution de ces groupes par la volonté de résoudre collectivement des problèmes[233]. Elle explique que les utilisateurs d'IRC mettent en place des mesures afin de résoudre leurs problèmes communs, sans système de régulation (modération), en dehors de tout contexte social physique, en ne disposant que d'une faible expressivité et dans un contexte d'anonymat. Les membres d'IRC font alors face à deux spécificités dans la constitution de leur communauté : l'absence d'expressivité non verbale et l'anonymat des personnes. Deux mesures ont été mises en place et participent du succès de la constitution de communautés (Reid, 1996). Premièrement, les membres « ont mis au point des systèmes de symbolisme et de signification textuelle » pour s'assurer de la compréhension de leur intention malgré l'absence de canaux de communication plus habituels. Il s'agit des émoticônes, aujourd'hui largement utilisées, et des « didascalies » qui complètent, comme au théâtre, l'intonation ou l'intention du message écrit. Deuxièmement, diverses sanctions sociales sont apparues au sein de la communauté IRC afin de sanctionner les utilisateurs qui ne respectent pas les règles du groupe et qui sont considérés comme nuisible à l'intégrité de la discussion (Reid, 1996). Michel Marcoccia observe également que l'écriture collective nécessite une gestion et une animation. En effet, le forum non modéré n'est pas un espace de publication dirigé par un éditeur[234] mais un cadre participatif plus proche de la conversation pour laquelle des règles sociales doivent s'appliquer (Marcoccia, 2001). Observant Usenet, il constate également que « les forums de discussion sont des espaces numériques de discussion qui permettent à des

[233] Ces problèmes sont souvent d'ordre informatique. On retrouve à nouveau le choix d'employer un nouveau système de communication dans la perspective de développer ce même système de communication.

[234] Ce n'est pas le cas des forums modérés, qui ne sont pas des cadres de production d'informations utiles à notre présentation sur la construction de communautés créatrices d'informations libres. Michel Marcoccia écrit à ce propos : « Au sens strict, lorsqu'un forum est modéré, les messages qui sont postés sont en fait adressés au modérateur qui choisit de les diffuser dans le forum ou de les exclure de la discussion. Dans ce cas, la modération du forum correspond clairement à une activité d'édition d'un document numérique dont la représentation affichée sur le support d'appropriation sera l'œuvre d'un éditeur particulier garantissant la cohérence du document par la sélection voire le 'reformatage' des messages adressés au forum » (Marcoccia, 2001).

utilisateurs de gérer des activités intellectuelles collectives, que ce soit de simples discussions ou des processus complexes de résolution de problèmes » (Marcoccia, 2001). À ce besoin de gestion de groupe, correspond l'apparition informelle du rôle d'animateur. Le membre animateur émerge comme coordinateur du groupe. Cet animateur peut, dans un processus d'organisation, se trouver attribuer le rôle formel de modérateur. Si ce modérateur possède des droits supérieurs aux autres membres, il n'est pas, dans les contextes d'échanges que nous évoquons, un intermédiaire entre l'écriture du message et sa publication. Son action éventuelle est postérieure à toute publication. Ainsi, Michel Marcoccia (2001) explique comment le groupe constitue progressivement des règles d'interactions. Il explique, citant Howard Rheingold[235], que des textes prescriptifs sont écrits « sur l'art d'animer une discussion en ligne ». On peut citer un certain nombre de règles qui se sont ancrées dans les pratiques comme la règle « Don't ask to ask, just ask »[236]. Plus hiérarchique encore que la description de Michel Marcoccia, Elizabeth Reid évoque les sanctions sociales qui s'appliquent lorsqu'un comportement « illégitime » est repéré par les membres de la communauté. Elizabeth Reid identifie trois sanctions sociales et structurelles :

- Le sentiment de honte suscité par et chez les membres de la communauté pour être allés à l'encontre du système de valeur élaboré par la communauté. Ce sentiment est relatif au degré de déconstruction des conventions sociales atteint au sein de la communauté et de participation à une communauté réglementée où la pratique de l'anonymat sur Internet ne doit pas apporter le plein pouvoir.
- La violence textuelle envers les membres considérés comme dans l'erreur.
- Le bannissement est également une pratique accessible aux utilisateurs dont le rôle est d'être « opérateurs » des forums IRC. En cas de bannissement, un message expliquant les causes de cette exclusion est envoyé au membre de la part de l'opérateur responsable.

Les communautés IRC sont fortement hiérarchisées autour d'un opérateur aux pouvoirs forts. Ainsi, alors que le média IRC présente une forte déconstruction des règles sociales, en mettant en contact des individus n'appartenant pas aux mêmes groupes sociaux, de nouvelles hiérarchies se constituent. Bien que non modérés, Usenet et IRC

[235] « Ainsi, Howard Rheingold (qui est un des promoteurs essentiels de la communauté virtuelle comme activité en ligne) définit le rôle de l'animateur en listant les objectifs suivants (Reinghold, 1998) : susciter des conversations « authentiques », faire naître un sentiment d'appartenance au groupe, générer de la créativité collective, favoriser l'autogestion des conflits, permettre les conditions de la collaboration entre membres du groupe, lutter contre la passivité des participants, accueillir les nouveaux arrivants. » (Michel Marcoccia, 2001).
[236] « Ne demande pas si tu peux demander, juste demande » (traduction par l'auteure).

sont des espaces sociaux, où des communautés se sont constituées, fortement réglementées et potentiellement fortement hiérarchisées.

Dans la suite de ces constructions techniques, une pratique de partage plus « grand public » que IRC et Usenet est apparue, c'est le *peer-to-peer* (pair-à-pair ou aussi P2P). Le ***peer-to-peer*** est une nouvelle forme de réseau sur internet. Reprenant les concepts initiaux de la forme de réseau 'client-serveur' comme le mail et le FTP, le *peer-to-peer* est cependant conçu comme un réseau décentralisé. Chaque ordinateur des utilisateurs du réseau est à la fois client (il peut recevoir des informations) et serveur (il peut diffuser des informations). En 2002, dans un article relevant de la sociologie du travail, Jean-Samuel Beuscart étudie Napster, le premier réseau de *peer-to-peer* tourné vers l'échange de fichiers musicaux. Il s'appuie sur des entretiens réalisés auprès d'utilisateurs français de Naptster ainsi que sur des matériaux complémentaires comme les forums et les listes de diffusions consacrés au thème du téléchargement musical. Ses travaux le mènent à qualifier Napster de dispositif socio-technique. En effet, le téléchargement de fichiers musicaux n'est pas apparu avec Napster (Beuscart, 2002).

IRC, Usenet et les mails étaient des espaces de rencontres virtuels entre internautes afin de s'échanger ces fichiers. Cependant, l'architecture technique était alors celle du client-serveur et les relations sociales étaient de l'ordre de la « négociation entre différents offreurs et demandeurs de fichiers musicaux » (Beuscart, 2002). La négociation repose, en effet, soit sur des mécanismes de réciprocité (troc) soit sur de la mise à disposition gratuite. Ainsi, en changeant de modèle d'infrastructure, « Napster détruit en grande partie la hiérarchie sociale qui s'était construite autour du contrôle des accès aux serveurs et affaiblit ou rend obsolètes des rituels d'échange contraignants » (Beuscart, 2002). L'innovation « organisationnelle » que représente Napster attire alors un large public et l'entreprise Napster Inc. présente ce public comme une communauté. Cette innovation organisationnelle mène donc à un changement d'échelle en termes d'utilisateurs. Jean-Samuel Beuscart se demande si Napster est un dispositif qui a permis l'émergence d'une communauté d'échanges auto-régulée dont les rapports sociaux sont fondés sur le don ou si Napster est un dispositif de consommation, régi par les comportements opportunistes, annonciateurs d'un marché de la musique en ligne. De cette possible ambivalence, Jean-Samuel Beuscart propose une synthèse. **Il insiste sur l'importance technique du dispositif qui a engendré une très forte fréquentation du réseau** (jusqu'à 800 000 connectés simultanément au réseau)

(Beuscart, 2002). Il explique que « la régulation du collectif relève de la solidarité technique » qui consiste en un partage des ressources des ordinateurs, de chacun, transformés en serveurs. Les aspects opportunistes et les actions morales de partage de l'information sont des éléments constitutifs du système Napster mais non fondamentaux et si l'ensemble du groupe ne constitue pas une communauté dont la visée serait le partage sans limite de la musique, Beuscart (2002) observe l'existence « de petites communautés spontanées au sein desquelles s'organisent des interactions réelles ». Ces communautés partagent d'abord des goûts musicaux et dans ce cadre échange des fichiers. Une forme de reconnaissance des individus émerge dans la reconnaissance du bon goût de certains diffuseurs. Ainsi, la structure de ces communautés est moins structurée que dans le cadre d'IRC.

4.1.1.3. D'un internet confidentiel au Web grand public

Entre le début des années 1970 et la fin des années 1990, l'accès à l'ordinateur personnel du grand public progresse. Si le modèle d'ordinateur Xeros Alto[237], diffusé en 1973, n'a pas atteint un large public mondial, Les ordinateurs personnels commencent à être diffusés à des particuliers non professionnels du domaine dans le courant des années 1980-1990. Au terme de cette période on peut citer le succès de l'iMac d'Apple sorti dans sa première version en 1998 qui touche un large public aux États-Unis et en Europe. Durant la période d'expansion, plus précisément, entre 1984 et 1991, les systèmes d'exploitation Macintosh (1984), Windows (1987) et Linux (1991) sont développés. C'est également au cours des deux décennies, de 1980 à 2000, qu'Internet devient un réseau mondial et sort des domaines universitaires et militaires. D'abord en 1983, le réseau Internet est ouvert. Puis, de 1989 à 1991, Tim Berners-Lee, ingénieur au Cern, conçoit, aidé de ses collègues du Cern, le World Wide Web. Il vise la **construction d'un système simple de partage de ressources qui exploiterait Internet tout en offrant un partage plus fluide**. Il présente ce système comme un espace auquel de nouveaux usages peuvent se greffer. Pour cela, son système doit imposer un minimum de contraintes techniques. À partir de 1990 et la publication d'une proposition de formalisation nouvelle comme structure de partage d'informations, sur Usenet, le Web est constitué de trois briques :

[237] Ce modèle d'ordinateur est parfois présenté comme l'un des premiers ordinateurs à écran et clavier dont l'usage peut être personnel comme dans la page Wikipédia qui le concerne (consulté en ligne : https://fr.wikipedia.org/wiki/Xerox_Alto).

- Le protocole HTTP (*HyperText Transfer Protocol*), protocole de requêtes vers un serveur web et de transferts de fichiers ou *ressources* (dans le langage du Web) depuis un serveur web ;
- Le formalisme HTML (*HyperText Mark-Up Language*) d'écriture de pages Web et de liens hypertextuels entre ces pages ;
- les URL/URI, des adresses uniques affectées à des ressources du Web et commençant souvent par « www ».

Une page HTML est un exemple de ressources disponibles via le Web lorsqu'une URL lui est affectée. Ainsi, pour qu'une ressource fasse partie du Web, elle doit être adressée par une URL ou une URI. Une adresse URL peut être atteinte en utilisant le protocole HTTP mais également le protocole FTP par exemple.

La page HTML est une publication statique. Cependant, son caractère principal est sa simplicité d'utilisation. Ainsi, couplés au protocole de requêtes et réponses qu'est le HTTP, des langages de scripts sont développés pour intégrer de l'interactivité à la navigation sur le Web. C'est par exemple le rôle du langage PHP[238], diffusé à partir de 1995. Un script PHP est exécuté après une interaction avec la page HTML par un utilisateur. Le script réalise une opération côté serveur et envoie une réponse au client. Ce procédé permet, de manière non technique pour l'utilisateur, de créer de l'information côté client, de l'enregistrer dans une base de données pour être potentiellement repris par un autre utilisateur par la suite. Des logiciels comme MediaWiki (logiciel actuellement utilisé pour l'ensemble des projets de la Wikimedia Foundation dont Wikipédia). Cependant, le langage PHP n'est pas l'unique langage adapté aux interactions serveurs-clients. Par exemple, c'est le langage Ruby qui est mobilisé pour le fonctionnement de l'API 0.6 d'OpenStreetMap dans ce but de rendre réalisable les interactions multiples, au travers d'une page HTML, entre les contributeurs et la une base de données enregistrées sur un serveur. Le réseau Internet reliant clients (logiciel utilisé par les contributeurs) et serveur (où est enregistrée la base de données OSM).

Au début des années 2000, le Web est donc un moyen simple pour se connecter à Internet. Plusieurs techniques d'échanges et d'interactivités ont déjà été développées et la constitution de communautés virtuelles se poursuit.

[238] PHP: Hypertext Preprocessor.

4.1.2. OSM, une construction socio-technique au sein d'Internet

Dans la section précédente, nous avons exposé les processus sociaux et les inventions techniques qui ont mené à la constitution de communautés d'internautes partageant des intérêts thématiques communs et parfois aussi des principes comme la réappropriation de l'organisation en communs. Nous poursuivons notre exploration des constructions socio-techniques ayant mené au projet OSM en nous concentrant cette fois sur sa production, une base de données géographiques. Dans cette section nous tentons, maintenant, de reconstituer les publications d'outils ou de recherches qui ont pu participer à l'émergence du projet OpenStreetMap afin d'en comprendre son fonctionnement. Nous sommes conscients des limites de l'exercice qui ne prend pas en compte les contextes personnels des quelques individus qui ont fait naître le projet OSM. En outre, notre éclairage met exclusivement en évidence les inventions techniques et les relations - commerciales ou du domaine de la recherche - qui sont encore opérationnelles aujourd'hui ou qui ont marqué le plus grand nombre. Nous laissons de côté certains co-développements entre logiciels et communautés ainsi que la question du choix des langages informatiques ayant permis d'écrire ces inventions, ce dernier étant à la fois susceptible de relever de choix personnels de leurs créateurs, de contraintes liées à la diffusion d'un produit ainsi que d'effets de mode en faveur de certaines technologies. Néanmoins, les éléments que nous avons retenus, en figure 66, sont, selon nous, aujourd'hui constitutif de l'écosystème de l'information géographique propriétaire ou libre, partagée *via* Internet ou commercialisée sur support papier ou numérique et constituent un socle stable construit par des experts.

4.1.2.1. Inventer les données spatiales et leurs outils

Nous commençons notre chronologie, dans les années 1960, avec le premier système d'information géographique connu, conçu au Canada par Roger Tomlinson. Docteur en géographie et pionnier de la cartographie assistée par ordinateur, Roger Tomlinson met en pratique ses recherches sur la structuration des données géographiques avec le *Department of Forestry and Rural Development* canadien pour l'aide à la décision de la gestion du territoire. Il publie en 1966 l'article « A Geographic Information System for Regional Planning ». L'article débute par le constat des limites auxquelles les gouvernements de l'époque font face pour administrer les ressources naturelles et l'occupation humaine de leur territoire. Il conçoit le système d'information

géographique comme la combinaison d'une « banque de données »[239] et d'un ensemble de méthodes permettant de manipuler et de comparer ces données[240] (Tomlinson, 1968). Nous retrouvons dans ce texte des éléments fondamentaux des SIG toujours en usage, comme, par exemple, la conception vectorielle des objets en points, lignes ou polygones. De même, des concepts de relations spatiales verticales, comme les intersections des objets, et horizontales, comme le calcul de distance, sont abordés. Dans le même temps, en 1965, à Harvard, le laboratoire *for Computer Graphics and Spatial Analysis* (LCGSA) est fondé. Ainsi, les chercheurs du laboratoire Howard Fisher expérimentent la transposition des concepts géographiques en programmes informatiques et parviennent à des conceptions proches de celles de Roger Tomlinson. Le laboratoire développe des programmes de cartographie automatique et de gestion de données géographiques. Lorsqu'en 1969, Jack Dangermond étudie au LCGSA, il bénéficie des programmes déjà existants pour proposer une analyse informatisée de la pollution de l'air à Los Angeles[241]. La même année il ouvre un bureau d'étude spécialisé dans l'exploitation de données géographiques environnementales qu'il nomme *Environmental Systems Research Institute* (ESRI). Mobilisant ses connaissances des recherches scientifiques dans le domaine de l'information géographique, Jack Dangermond développe avec son entreprise un ensemble technique qui transpose les concepts de l'analyse spatiale en méthodes informatiques. En 1990, le logiciel ARC/INFO diffusé par ESRI depuis huit ans, est le SIG commercial le plus diffusé (Peuquet et Marble, 1990). Le logiciel, destiné à des stations individuelles de travail (PC), est entièrement placé sous licence propriétaire.

Au début des années 1990, le langage de gestion de données relationnelles SQL[242] est intégré au logiciel ARC/INFO et les recherches contemporaines se tournent vers la définition d'un système de requêtes spatiales en utilisant le SQL. Dans le rapport *Towards SQL Database Extensions for Geographic Information Systems* (1993), dirigé par Vincent B. Robinson de l'université de Toronto et Henry Tom du NIST (*National Institute of Standards and Technology*) aux États-Unis, chercheurs et acteurs

[239] « *data bank* » (Tomlinson, 1968).
[240] « *The set of procedures and methods for moving data into the bank, and for carrying out the manipulations, measurements and comparisons of the data, once there* ».
[241] Article de célébration des cinquante ans du LCGSA sur le site d'ESRI : https://www.esri.com/esri-news/arcnews/winter1415articles/celebrating-the-advent-of-digital-mapping (consulté en ligne en avril 2019)
[242] Le SQL est alors un standard ISO depuis 1987, selon la publication *ISO 9075:1987 Information processing systems -- Database language -- SQL*

économiques, membre du groupe de travail *GIS/SQL* sont associés afin de proposer un standard pour la manipulation de données spatiales. Parmi les acteurs économiques nous trouvons, Jon W. van Roessel, de l'entreprise ESRI. Il évoque les concepts nécessaires à la conception de jointures spatiales entre couches de données vectorielles décrites en points, en lignes ou en polygones (p.133). Nous trouvons, dans son texte, les éléments des outils aujourd'hui largement utilisés au travers de la suite ESRI, que soit les intersections (*intersect*), les superpositions (*overlay*) ou les unions (p.138). Bien qu'il existe des partenariats entre laboratoires de recherche et entreprises, l'essentiel des applications est protégé d'une licence propriétaire. Ce n'est cependant pas le cas du logiciel libre de SIG, GRASS GIS (*Geographic Resources Analysis Support System*). La création de ce logiciel remonte au tout début des années 1980, au sein du CERL (*Army Corps of Engineer's ConstructionEngineering Research Laboratory*) en Illinois. Westervelt (2004) : un de ses premiers concepteurs et étudiants à l'époque, relate ces débuts en miroir du contexte général des SIG de l'époque. Il affirme que le SIG ARC/INFO est, en 1980, le SIG le plus performant au monde. Le SIG était cependant payant et même onéreux puisque, le logiciel coûte alors 15 000 dollars et exige l'utilisation d'un mini-ordinateur d'une valeur de 60 000 dollars. Nous comprenons alors que les motivations qui ont mené à la conception de GRASS GIS correspondent à la situation récurrente des motivations du développement d'une solution libre. Les chercheurs du CERL décident de développer leurs propres solutions logicielles adaptées à leurs moyens et notamment à leurs faibles ressources matérielles. Ce sont ces conditions de développement qui ont mené à la conception d'un SIG composés en modules. Ces modules sont des fonctions informatiques répondant à un besoin spécifique de gestion de données spatiales. Les ressources matérielles (la mémoire et le processeur) ne sont mobilisées qu'au lancement d'un module par l'utilisateur. Cherchant à expliquer le succès de GRASS, Westervelt (2004) résume les originalités du projet, qui en font, selon nous, un modèle du logiciel libre, développé malgré lui :

- Un puissant SIG placé sous les conditions de diffusion du domaine public, nécessaire pour soutenir la planification environnementale.
- Une alternative peu exigeante en termes de coûts en matériel, ce choix de développement est adapté aux modestes ressources des premiers utilisateurs.
- Un système ouvert de programmation, permettant le développement efficace de nouvelles méthodes d'analyse par SIG.

- Le système principal (bibliothèques ou API) était géré par un groupe restreint, mais toute personne intéressée pouvait améliorer les programmes existants ou en créer de nouveaux.
- Des réunions annuelles d'utilisateurs étaient organisées pour aider à maintenir et développer une communauté autonome.
- Un bulletin d'information permet aux utilisateurs de prendre connaissance des nouvelles opportunités, les expériences des autres et le support[243].

Nous retrouvons dans ce résumé les principes d'un développement ouvert favorisant l'itération au sein d'une communauté qui communique régulièrement et qui place ses productions au profit du plus grand nombre et notamment de la recherche et des services publics. Si GRASS vise, dès sa conception, un public aux moyens limités, il s'adresse à des experts du développement et de l'informatique. L'ensemble des modules est mobilisé essentiellement par le moyen de lignes de commande. Une telle pratique nécessite assurément une expérience de l'informatique.

À cette limite, Gary Sherman répond par la publication du logiciel SIG QGIS (Quantum GIS jusqu'en 2013). De deux contributeurs la première année, le projet réunit en 2019 plus de cinquante contributeurs, différents selon les mois[244]. Sur le site de présentation et de documentation du logiciel, on peut lire, « QGIS est un Système d'Information Géographique (SIG) **convivial** distribué sous licence publique générale GNU. C'est un projet officiel de la fondation Open Source Geospatial (OSGeo). Il est **compatible** avec Linux, Unix, Mac OS X, Windows et Android et **intègre de nombreux formats** vecteur, raster, base de données et fonctionnalités ». Présentant une interface graphique, le logiciel peut intégrer les modules de GRASS, sur ses fonctionnalités et partage encore aujourd'hui le développement de codes sources. Il s'adresse à un large public, de l'amateur au développeur, adepte des solutions libres aux utilisateurs de systèmes d'exploitation et de formats de données propriétaires. **Il s'agit encore d'un trait caractéristique du mouvement du Libre, qui, tout en se positionnant en alternative des solutions propriétaires intègre leurs standards, comme, par exemple, le Shapefile d'ESRI. Le développement d'outils libres vise l'interopérabilité pour des usages diversifiés et s'adapte donc à leurs cadres**, comme, par exemple, QGIS

[243] Traduction par l'auteure.
[244] Les informations relatives aux contributeurs sur le dépôt de code Github sont disponible à l'onglet *Insights* : https://github.com/qgis/QGIS/graphs/contributors (consulté en mai 2019)

pouvant être installé sur les systèmes d'exploitation de Microsoft et d'Apple. Dès 2002, date de sa première version, l'import et la visualisation de données enregistrées sous le format PostGIS deviennent possibles. PostGIS est une extension du SGBD PostgreSQL, lui apportant des fonctionnalités propres à la gestion de données spatiales. L'article « *PostGIS History* » publié sur le site http://www.refractions.net/ résume les étapes de son développement. Encore aujourd'hui largement développé par l'entreprise canadienne Refractions Research, PostGIS est le résultat d'une demande d'un ministère du gouvernement de la province canadienne Colombie britannique. En 2000, utilisateur du SGBD PostgreSQL[245], Refraction Research travaille, pour ce ministère, à la mise en place et au fonctionnement de systèmes de gestions de données. Le SGBD PostrgeSQL bien que proposant des fonctionnalités de gestion de données spatiales est jugé insuffisamment performant. Profitant de ses connaissances du SGBD PostgreSQL et de sa nature ouverte, l'entreprise décide de construire des nouvelles fonctionnalités à partir de ce logiciel hôte. L'extension PostGIS est aujourd'hui mobilisable au travers de logiciels tels qu'ArcGIS[246], GeoServer[247], Grass GIS, MapInfo[248], Mapnik[249], MapServer[250], ou QGIS.

Cette interopérabilité avec des outils aux fonctions et aux utilisateurs diversifiés, s'explique notamment par l'intégration de l'extension PostGIS à un ensemble technique cohérent coordonné par la Fondation Open Source Geospatial (OSGeo). Fondée en 2006 par Fondée en 2006 par des groupes de contributeurs à des projets ouverts (dont GRASS, GDAL/OGR et Geoserver), l'OSGeo s'est donné pour rôle de fournir un soutien financier, organisationnel et juridique à la communauté de contributeurs aux outils de gestion de données spatiales *open source*. Pour les projets de développement, l'OSGeo représente une institution garante d'une protection juridique et facilitatrice pour l'interopérabilité entre les projets qu'elle défend. Nous retrouvons ces interactions entre les projets soutenus par l'OSGeo dans notre sélection d'outils et d'acteurs constitutifs de l'écosystème d'OSM (figure 64). Ainsi, sont fédérés au sein de l'OSGeo

[245] Le SGBD PostgreSQL est un des résultats de vingt ans de recherche menés par le chercheur en informatique, Michael Stonebraker, titulaire du prix Turing 2014 pour ces travaux. Le SGBD est publié, hors d'un serveur inter-universitaire, en 1996 en *open source* par Bruce Momjian (German, 2006).
[246] Logiciel SIG de l'entreprise ESRI.
[247] Logiciel serveur de données géographiques répondant à des requêtes par les clients pour le partage des données et leurs modifications en ligne.
[248] Logiciel SIG de l'entreprise Pitney Bowes.
[249] Logiciel libre de service de tuiles cartographiques. Il sert à la création de la *slippy map* sur openstreetmap.org.
[250] Plateforme libre en ligne pour la publication de données spatiales et la construction d'applications utilisant des données spatiales.

les bibliothèques de développement GDAL/OGR et PROJ.4, l'extension PostGIS du SGBD PostgreSQL, les services de rendus cartographiques en ligne MapServer et OpenLayers, les logiciels QGIS et GRASS GIS. Assurant un système de contrôle de qualité lors de l'intégration de chaque projet communautaire à la Fondation, l'OSGeo assure un rôle de garante vis-à-vis des utilisateurs. L'OSGeo est ainsi une sphère assurant auprès des utilisateurs, et notamment des utilisateurs-investisseurs, la pérennité du projet, mais aussi son ouverture et son interopérabilité.

4.1.2.2. Les données spatiales en ligne

À la fin des années 1990, a été produit un ensemble de concepts et de réalisations techniques ouvertes. Le but est d'assurer la diffusion et l'usage de l'information géographique par un public plus large que le public universitaire. Portée par la volonté du mouvement du Libre de diffuser les connaissances, cette propagation des outils désormais « conviviaux », selon l'expression de la description de QGIS, se fait par le Web. A ce stade de l'évolution, nous avons atteint la période du Web interactif, discutée dans la section 1.1 (l'interaction est techniquement permise par les langages PHP et JavaScript, par exemple), de partage et d'écriture de documents en ligne. Les moyens techniques de diffusion en ligne des images fixes (sous les formats JPEG et PNG) et de fichiers de formes (sous le format SVG) se standardisent également à la fin des années 1990. En mettant en relation les concepts qui ont servi à l'informatisation de l'information géographique et la diffusion d'images, la cartographie commence à de devenir une pratique de communication sur Internet. En 2000, mobilisant les bases de données relationnelles, la bibliothèque de conversion de projections cartographiques PROJ.4 et les standards d'écriture des images fixes et vectorielles, l'Open Geospatial Consortium (OGC) propose un nouveau standard, le *Web Map Service* (WMS). Le service Web de carte met en relation un utilisateur appelant depuis un navigateur un document – le plus souvent sous format PNG et JPEG – géo-référencé. Ce document est construit par un serveur cartographique de tuiles (cf. chapitre 3), lui-même connecté à une base de données spatiale. Les sites de consultation de cartes tels que le Géoportail, Google Maps et OpenStreetMap sont les applications directes du procédé WMS proposé par l'OGC.

Ces nouvelles représentations cartographiques proposent peu à peu une couverture mondiale (bien que non homogène localement en ce qui concerne la qualité des

données) et des outils de localisation des informations menant au Géoweb. Deux services viennent compléter ces cartes en ligne et de plus en plus interactives : Google Earth et Google Street View. La diffusion du savoir géographique atteint le public étendu des internautes qui, progressivement, en deviennent des producteurs. L'information spatialisée n'est plus uniquement produite dans les contextes clos des Universités, des organismes de recherche, des instituts de géographes professionnels, des États, des ONG, des militaires, des producteurs de cartes et d'imagerie satellite (Bakis et Valentin, 2010). **Le Géoweb est le contexte de production d'informations souvent très élémentaires mais parfois aussi approfondies dans leurs méthodes et dans leurs contenus**. Le concept de néogéographie suit presque simultanément ces premières pratiques de production de données spatiales en ligne par le moyen de services grands publics comme Google Maps et ses premiers *mashups*. Apparu entre 2004 et 2006, pour sa définition la plus moderne, relative à des pratiques spécifiques sur le Web, le terme se réfère à une tendance répandue sur le Web. Elle consiste à créer et à utiliser des outils présentant des fonds de carte et permettant des interactions avec ce fond de carte. Mais le concept de néogréographie suppose que les utilisateurs sont des « amateurs ». Cette caractéristique de « non professionnel » nous semble si restrictive, que nous excluons de la néogéographie les pratiques collectives organisées décrites dans cette thèse. Alors que les techniques du Web permettent à un public assez large l'annotation et la géolocalisation de points sur des fonds de carte (comme ceux de Waze et Foursquare) et éventuellement la conception d'une couche thématique d'informations à ajouter sur un fond de carte (le *mashup* pratiqué par des associations ou des particuliers intéressés par une thématique), d'autres usages de ces techniques nous semblent accessibles seulement à un public plus restreint. En effet, comme nous l'avons déjà vu, l'élaboration d'infrastructures techniques et de méthodes collectives de création d'informations géographiques a attiré un public expert de la géomatique, de la géographie et de l'informatique et a conduit des contributeurs, potentiellement néophytes, mais devenus des *senior mappers*, à acquérir de nouvelles compétences techniques et méthodologiques. Pour contextualiser le projet original d'OSM ainsi que les outils et les constructions de services créés à partir d'OSM, nous préférons donc employer le terme de Géoweb, moins excluant dans ses acteurs et davantage tourné vers la construction d'un système d'informations en ligne que le terme de néogéographie.

Des années 1990 aux années 2010, Internet a été mobilisé par une population élargie, aux besoins diversifiés. Au cours de cette période, le Web interactif se développe grâce à des techniques permettant soit des échanges multilatéraux entre le serveur et le client (PHP, Ajax, Ruby), soit la possibilité, pour l'utilisateur d'un logiciel Web-client, de modifier le contenu de la page sans faire appel aux contenus enregistrés sur un serveur (Flash, JavaScript). Ces techniques, d'abord mobilisées sur les ordinateurs, sont depuis les années 2010 mobilisables depuis les téléphones portables, dits *smartphones*. Leur rôle dans le développement du Géoweb n'est pas anecdotique. Les divers services sont largement consultés depuis les smartphones, notamment les calculs d'itinéraires. En outre, la pratique de contribution la plus simple, qui consiste à noter des lieux ou à annoter des points d'intérêts est réalisée depuis ces mêmes terminaux. Ces avancées techniques sont motrices pour le développement d'une expression populaire en ligne, par le moyen des cartes interactives. Lorsqu'en 2010, Henry Bakis et Jérémie Valentin se demandent si « la consommation de l'espace deviendra augmentée »[251] et si l' « usage espace/cyberespace préfigure des changements de comportement face, par exemple, aux mobilités urbaines et touristiques », nous pouvons désormais répondre que leurs prévisions se sont réalisées (fig. 67 et fig. 68). Ainsi, en 2018, Renaud Le Goix écrit-il : « dans le monde contemporain, l'information géographique est partout, ubiquiste, aisément disponible. Qu'il s'agisse de la cartographie sur « smartphone » pour se déplacer ou se repérer, de la géolocalisation des vélos ou voitures en libre-service, de la recherche d'itinéraires optimisés voir de la rencontre d'amis ou de partenaire dans le voisinage proche, les utilisations sont quotidiennes et chacun peut appréhender l'espace ».

[251] Ou réalité augmentée qui consiste à l'ajout d'éléments d'informations à ce qui est observable au travers d'un terminal et notamment d'un téléphone portable.

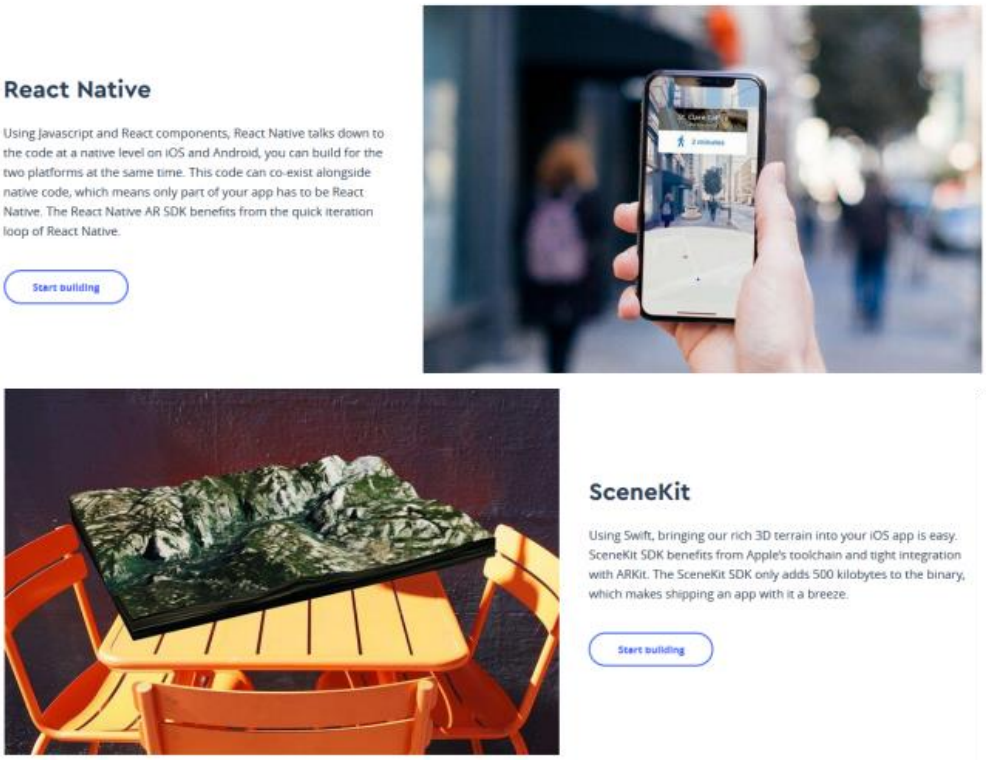

Figure 67 Services commerciaux de réalités augmentées à partir des données OSM et d'images satellites proposés par Mapbox[252]

Figure 68 Présentation par Waze de son utilisation des données fournies par ses utilisateurs[253]

[252] Sources : Captures d'écran de la page *augmented reality* du site Web de Mapbox, consulté en ligne en mai 2019 : https://www.mapbox.com/augmented-reality/
[253] Source : Capture d'écran d'une partie de la page d'accueil du site Web de Waze, consulté en ligne en mai 2019 : https://www.waze.com/fr/

4.2. Production par les pairs

Du web interactif émergent deux pratiques de production de connaissances, le pair à pair (*peer to peer* sans vouloir faire référence au procédé technique et numérique*)* et le *crowdsourcing*. Le procédé d'échanges de pair à pair a été illustré précédemment par l'exemple de Napster, plateforme d'échanges de pair à pair en ligne, que nous présentons comme une invention représentative du processus de mobilisation d'innovations techniques par des groupes partageant des intérêts communs (voir la section IV.1.1.2. présentant les protocoles techniques et sociaux de communications). Par ailleurs, nous déjà avons abordé la pratique du *crowdsourcing* dans notre recherche de définition du projet OSM (voir la section I.2.2.Volontaire, contributif). À la définition du *crowdsourcing* par Jeff Howe proposée en 2006, qui a popularisé le concept, nous préférons celle d'Enrique Estellés-Arolas et de Fernando González-Ladrón-de-Guevara datant de 2012. Ceux-ci définissent le *crowdsourcing* comme l'appel à une foule lancé par un individu (le plus souvent une institution ou une entreprise) qui soumet un problème à résoudre. La résolution du problème se pratique en ligne à travers 'une plateforme. La pratique de production de connaissances est donc dirigée par une demande et sa réalisation est rétribuée. Comme le souligne Théo Zimmermann, pour le cas de Wikipédia : « l'encyclopédie ne tient pas dans les limites restrictives du *crowdsourcing*. En effet, personne en particulier n'est à l'origine de l'appel à la foule. Quant à la tâche, bien définie, de construire une encyclopédie, elle est sans fin » (dans Orsi et *al.*, 2017). Ces observations s'appliquent de la même manière au projet OSM, dont la tâche est de maintenir à jour une base de données géographique prenant en compte toute la richesse des territoires. **Openstreetmap ne relève donc pas véritablement de la pratique du *crowdsourcing*.**

Cette sous-partie interroge la possibilité de qualifier OSM de système de production entre pairs. **Dans quelle mesure le projet OSM est-il un système de production décentralisé, dont l'objectif est le reversement du résultat obtenu dans un commun[254] ?**

Comme nous avons commencé à l'aborder dans le troisième chapitre, la question de l'exploitation d'une main-d'œuvre par une entreprise qui mobilise un grand nombre

[254] Ces caractéristiques d'une production pair à pair sont tirées de la définition de Théo Zimmermann dans Orsi et *al.*, 2017.

d'internautes, qu'ils soient volontaires ou non, relève d'une nouvelle problématique issue d'Internet à laquelle les contributeurs à OSM doivent se confronter, afin de se prémunir du risque de devenir une plateforme de *crowdsourcing*.

Nous considérons que le procédé de collecte des connaissances par le *crowdsourcing* représente un risque pour le projet OSM, au sens où il contrevient aux principes de liberté de contribution et à une organisation autogérée dans ses objectifs comme dans ses procédés techniques. Le *crowdsourcing* est un procédé controversé lorsqu'il prend la forme d'une activité précaire pour des participants rémunérés à la tâche, dont la production profite exclusivement au commanditaire lequel a ainsi réduit ses coûts de main d'œuvre.

Cependant, de telles conséquences négatives correspondent à une forme réductrice d'application du principe consistant à proposer un problème à une foule. Par son ampleur, une foule peut répondre de manière créative et originale à une question. En outre, la réponse peut être partagée publiquement. C'est ainsi que la recherche a parfois recours à la production par *crowdsourcing* : ce sont les sciences participatives. Les projets de sciences participatives sont menés par des institutions de recherche qui invitent les citoyens à contribuer à l'amélioration des connaissances sur des sujets comme la biodiversité[255], l'astronomie[256] et la biochimie[257]. En France, le Muséum National d'Histoire Naturelle (MNHN) est particulièrement actif dans l'exploitation des capacités des citoyens à participer au développement des connaissances scientifiques. Appelant à une participation active et délibérée, le MNHN présente ses projets de sciences participatives sur la biodiversité comme des solutions pour des chercheurs aux moyens limités. Pour une organisation fine de la collecte, le MNHN s'appuie, en autres, sur le groupe de contributeurs de Tela Botanica, spécialisés dans le domaine botanique. Les partenariats entre l'association et le centre de recherche sont fréquents. Nous signalons ainsi, par exemple, le programme de science participative « Sauvages de ma rue » qui est à la fois, un projet pédagogique animé par l'association Tela Botanica et un projet scientifique du laboratoire CERSP du MNHN. Les données collectées sont ensuite exploitées par l'association et par les laboratoires afin de publier

[255] Comme le réseau d'observatoires Vigi-Nature de suivi des espèces en France : http://www.vigienature.fr/

[256] Comme le projet zooniverse, largement tourné vers l'exploitation de document astronomique : https://www.zooniverse.org/

[257] Le jeu *Foldit* propose aux utilisateurs de mettre à profit leur logique pour la compréhension des formes que peuvent prendre les molécules. Le jeu consiste en la résolution de puzzles afin de transférer la méthode instinctive de l'humain à un algorithme.

des travaux à visées pédagogiques[258] et scientifique. Il s'agit là d'un partenariat intéressant entre un projet de collecte par *crowdsourcing* et un projet de conception et de collecte entre pairs.

Afin d'estimer si le projet OSM est un système de production entre pairs, c'est-à-dire décentralisé mais néanmoins coordonné autour d'un commun, nous revenons sur la définition précise du commun informationnel (Coriat, 2015) et nous exposons dans quelle mesure OSM répond aux principes des communs et en quoi il s'en détache.

4.2.1. Communs informationnels

Dans le premier chapitre, nous avons défini la communauté OpenStreetMap et les communautés du Libre comme des productrices de *communs*. Les communs sont des ressources partagées par un groupe de personnes. Ces ressources présentent des contours définis par le groupe, de manière à assurer une production et une exploitation raisonnée de la ressource afin de garantir sa pérennité. Nous mobilisons le concept de *communs* au sens défini par Elinor Ostrom (Ostrom, 1990). Celle-ci ne conçoit pas davantage de caractéristiques élémentaires aux *communs* que ce que nous avons explicité ici. Néanmoins, la chercheuse arrive à la conclusion que la pérennité d'un *commun* est garantie lorsque le partage est encadré par un ensemble de règles. Si elle propose huit principes généraux d'agencement en communs favorables à la pérennité d'une ressource partagée par un groupe, Elinor Ostrom souligne que l'originalité des règles concrètes choisies par les groupes qu'elle a observés est un facteur de longévité. Partant de ces conclusions, il nous importe maintenant de comparer l'agencement général du projet OSM à celui du ou des communs théoriques d'Ostrom. Nous comparons certaines pratiques qui se déroulent au sein de *communs* numériques que nous avons observés dans cette thèse (Wikipédia, Wikidata, les logiciels libres, mais surtout OpenStreetMap) par rapport au concept développé par Elinor Ostrom en n'omettant pas la spécificité du contexte d'Internet.

La base de données OpenStreetMap est un bien commun informationnel dont la ressource est la base de données OSM. Cette ressource informationnelle présente des

[258] En 2012, un article à destination d'un jeune public résume les premiers résultats issus de Sauvages de ma Rue : https://www.vigienature-ecole.fr/les-observatoires/comment-participer-sauvages-de-ma-rue/resultats-scientifiques-de-sauvages-de-ma

différences de plusieurs ordres avec une ressource classique d'un commun tels qu'un moulin, une rivière ou encore une zone de pêche. Ces différences mènent nécessairement à une adaptation, parfois inconsciente[259], des principes construits par les groupes partageant des communs physiques. Les biens communs informationnels sont (selon la définition de Benjamin Coriat dans Orsi et *al.*, 2017) :

- non rivaux et non-exclusifs,
- protégés par des droits de propriété intellectuelle innovants,
- partagés par un groupe dont la visée est l'enrichissement de la ressource davantage que la conservation des connaissances.

Les biens non-rivaux sont des biens qui ne peuvent s'épuiser. C'est le propre du document numérique qui peut être copié sans limite. La qualité de non-exclusivité correspond à cette possibilité technique de copier un document numérique, de manière à ce que le document puisse être à la disposition d'un individu sans en priver un autre. Les caractères non-rivaux et non-exclusifs des informations répondent aux principes de production entre pairs posés par la science moderne et repris dans la pratique de productions libres comme le rappelle Hess et Ostrom (2007, p.5). Elles soulignent que le partage avec le plus grand nombre est un élément de l'enrichissement de la ressource. Elles expliquent ainsi, sans le citer, le procédé de production dit itératif, ou dit cumulatif, selon lequel plus il y a de personnes qui partagent des connaissances utiles, plus le bien commun croît. Cette première spécificité du bien commun de type informationnel a pour corollaire son mode de préservation original.

Comme nous l'avons expliqué dans le premier chapitre, **cette préservation est issue de l'émergence rapide, au cours des années 1980, d'un marché des connaissances et des compétences informatiques**. Les communautés organisées autour d'un commun informationnel se sont constituées en ligne, partageant sans limite technique (l'exigence d'un mot de passe par exemple) ou juridique (celle d'une licence) leurs productions. Alors que la pratique depuis les années 1960 était, entre chercheurs, du partage de pair à pair concentré sur le partage plutôt que sur le brevetage, le rendement économique est finalement apparu comme une autre voie nécessitant la restriction de l'accès aux connaissances. Une tendance à la fermeture des codes alors publiés sans copyright a

[259] Les travaux d'Ostrom (1990) et Coriat (2015) soulignent la proximité des mécanismes entre les communs tangibles et les principes du Libre tout en rappelant que les seconds ont été conçus par des personnes, pour lesquels le concept de communs était inconnu, en réaction à un phénomène d'enclosure des connaissances.

abouti à un conflit éthique et politique à propos des connaissances numériques. Nous nous situons là au début de la commercialisation des connaissances informatisées (et notamment des logiciels), dans les années 1980. Plus de trente ans après la première licence *copyleft*[260] , la licence ODbL est aujourd'hui une licence principalement conçue pour protéger efficacement et simplement la base de données OSM et reprise par certaines institutions contributrices au modèle ouvert. La licence est à la fois une protection et un manifeste politique et éthique. Si les règles de protection juridique libres ont aujourd'hui atteint une certaine maturité, le « droit de prélèvement » est, dans le cadre de projets libres comme OpenStreetMap ou Wikipédia, défini par des formulations simplifiées (Coriat dans Orsi et *al.*, 2017). Comme nous l'avons discuté dans les premier et deuxième chapitres, les contributeurs à OpenStreetMap, présentent un intérêt pour la licence et débattent de son respect lors d'imports de données et d'utilisations commerciales. Néanmoins, cette protection n'est pas centrale dans l'activité des contributeurs. L'activité principale du producteur consiste en effet dans la création de contenus.

Cette tendance prédominante à la production et à l'enrichissement est considérée par Coriat comme le troisième élément définissant un commun informationnel. La production n'est pas distincte de la notion de prélèvement de ressources. Coriat explique « qu'à côté des règles qui régissent le prélèvement de ressources [la licence ODbL, dans le cas d'OSM], le commun informationnel se distingue par le fait qu'il comporte des séries de règles précises qui ont trait aux conditions de l'enrichissement du commun en informations et connaissances nouvelles [le *wiki* OSM] » (Coriat dans Orsi et *al.*, 2017). En effet, la licence *copyleft*, catégorie de licence libre à laquelle la licence ODbL appartient, est conçue pour partager la ressource et assurer son enrichissement, qu'il soit ou non réalisé dans le contexte principal du commun. Ainsi, comme dans le contexte de la recherche scientifique ouverte, lorsqu'un individu produit de nouvelles informations, à partir des informations publiées sous une licence *copyleft* et lorsque cet individu utilise publiquement ces nouveautés, il est tenu de les rendre disponibles au commun initial. Coriat (dans Orsi et *al.*, 2017) cite deux types de règles, les règles légales de protection et les règles de conditions d'enrichissement. Le *wiki* du projet OSM présente donc une richesse toujours croissante. Ses règles de conditions d'enrichissement, actées par consensus, visent une production collective efficace.

[260] La licence GNU-GPL définie entre 1983 et 1989.

L'analyse que nous avons proposée de la construction des règles de contribution ainsi que celle relative aux licences ODbL et Creative Commons, nous semblent recouper les conclusions de Benjamin Coriat. La protection de la propriété intellectuelle des communautés est une problématique centrale mais elle est largement dépassée par celle du partage de conventions de contributions, qui garantit l'enrichissement de la ressource.

Ces caractéristiques du commun informationnel montrent que l'aspect social est central dans le projet OSM. Ainsi que le définit Elinor Ostrom, le commun consiste en un système de partage entre individus. Les règles qui régissent ce partage exigent des individus qu'ils entrent en contact par l'échange de documents numériques. Ces documents soutiennent l'enrichissement de la ressource, par la communication de documents légaux et la mise en place de moyens de repartage, eux aussi au service de l'enrichissement, et par l'ensemble des méthodes établies afin d'enrichir utilement la ressource.

Chacun des exemples de communs numériques (OSM, Wikipédia, Wikidata, Tela Botanica) que nous avons proposé, exploite les techniques numériques et les interactions sociales qui s'établissent via **le réseau Internet**. Dans ce chapitre, nous avons exposé dans quelle mesure Internet a lui-même été pensé en commun par des chercheurs et des acteurs économiques comme un réseau ouvert, durant des échanges horizontaux informels ou des communications scientifiques, favorisant les échanges entre pairs et soutenant la liberté d'expression. Nous avons montré l'importance du contexte socio-technique initial (depuis les projets DARPA et CYCLADES) pour l'invention continue de techniques et protocoles standardisés internationalement permettant toujours davantage de partage des connaissances. Le protocole FTP, le *mail*, le protocole TCP/IP sont, par exemple, des applications techniques de concepts informatiques aujourd'hui mis en communs au service du partage d'informations dans le monde qui participent de la constitution de communs numériques. En outre, ces nouveaux moyens techniques de communication servent de lieux d'échanges pour l'invention de nouvelles techniques et la constitution de nouvelles communautés scientifiques ou hors des institutions traditionnelles. Nous citons ces trois inventions à dessein en tant que moyens de communication entre individus et de mises en partage de documents écrits puis, avec l'invention de formats numériques divers, de musiques, d'images et de vidéos.

Si nous avons examiné un système socio-technique et avons jusqu'à présent insisté principalement sur son aspect technique, nous n'ignorons pas que le développement de ce système tient également à des interactions sociales resserrées. à ce sujet, en 2012, Peter Mooney et Padraig Corcoran, travaillant sur la région de Londres, posent la question : « *How social is OpenStreetMap* ? »[261]. Ces deux chercheurs ne disposent d'aucune trace en ligne permettant de mesurer les interactions sociales entre les contributeurs actifs Ils tentent cependant d'exposer ces interactions au travers de potentiels coéditions d'objets. Ils observent qu'un tiers des objets ne dépassent pas deux versions. Il s'agit pour eux d'une première preuve de l'isolement des contributeurs londoniens. Finalement, leurs conclusions envisagent des types de contributeurs qui seraient soit des créateurs d'objets, soit des éditeurs des données déjà existantes, soit des contributeurs plus généralistes.

En complément de ces premières observations qui exposent les faibles interactions sociales au moment de la contribution à la base de données, nous souhaitons mettre en évidence le rôle des rencontres physiques dans la constitution de sous-communautés d'OSM. Pour cela, nous utilisons la liste des événements déclarés sur le *wiki* OSM (*Current events*, 2018) ; toute personne inscrite sur le *wiki* est autorisée à renseigner ces informations. Nous comptons 540 événements signalés pour l'année 2015, 620 en 2016, 640 en 2017, 500 en 2018. En mai 2019, 310 événements étaient déjà signalés. Ces chiffres montrent que l'activité sociale lors de rencontres physiques est conséquente et croissante au travers du projet OSM. Par ailleurs, les événements présentent des formes très différentes. Dans certaines villes, les contributeurs, suffisamment nombreux, organisent des événements réguliers durant lesquels ils se réunissent pour un moment de convivialité. Ainsi, des rencontres ont lieu à Rennes, Montpellier, Stuttgart, Bruxelles, Edinburgh ou encore Buenos Aires. D'autres réunions sont organisées afin de contribuer collectivement à OSM, comme Les Mercredis cartographie à Montréal et les promenades de collectes de données (ou *mapping-parties*). On peut également relever l'organisation d'ateliers comme l'*OpenStreetMap Workshop with Women and for Women* à Quezón City aux Philippines, *Las calles de las mujeres: OpenStreetMap + Wikidata + Wikipedia* à Córdoba en Argentine ou encore les *Mapathon*, organisés par l'association *Humanitarian OpenStreetMap Team* notamment. Ces ateliers sont davantage ouverts à un public extérieur mais concerné. Enfin, des réunions à caractère

[261] Dans quelle mesure OpenStreetMap est-il social ?

plus officiel, organisées en conférences, se tiennent également sous le nom *State Of The Map* (SOTM). Ces réunions, d'abord internationales (chaque année depuis 2007, à l'exception de 2015) puis régionales (SOTM Afrique, SOTM Amérique Latine et SOTM Europe) et nationales (SOTM France, SOTM Espagne), réunissent jusqu'à plusieurs centaines de personnes intéressées par le sujet du projet OSM. Les SOTM sont des conférences de plusieurs jours où l'on échange à propos de nouvelles techniques développées et où le devenir du projet est débattu au regard de l'état de la carte. Sur le même format, nous relevons que les conférences FOSS4G[262] sont signalées sur le *wiki*. En effet, certains contributeurs à OSM, partagent des intérêts communs. Ils aspirent au développement d'outils *open source* pour la géographie et partagent ces intérêts lors de rencontres formelles ou informelles hors des moyens numériques.

4.2.2. OSM : une organisation autour d'un bien commun façon Ostrom ?

Dans cette dernière sous-partie, nous posons la question de la réalisation du projet OpenStreetMap en tant que commun et des perspectives pour le projet : est-il possible d'envisager OSM comme une organisation horizontale, ouverte et inclusive ? En effet, l'accroissement de la ressource qu'est la base de données OSM ainsi que le développement de la communauté par le nombre de contributeurs mais également par l'exploitation commerciale de la base de données, ont pour enjeu la continuité du projet OSM. Les travaux de définition des concepts du domaine des communs d'Elinor Ostrom et Benjamin Coriat n'intègrent pas la question de la taille des communautés comme une limite pour leur pérennité. Il nous semble néanmoins que la croissance rapide et l'intégration d'acteurs privés et publics constituent un défi pour la communauté. Nous présentons ici deux exemples des conséquences de ces rapides transformations.

Le premier exemple est développé par Maude Gauthier et Kim Sawchuk (2018), chercheures, spécialistes des médias, qui relatent leur pratique de contribution pour la vulgarisation de leurs recherches sur Wikipédia. Ces chercheuses contribuent à Wikipédia sur la thématique du vieillissement afin d'inclure les études critiques et culturelles relatives au vieillissement. Elles choisissent cette thématique en raison de sa

[262] La *Free and Open Source Software for Geospatial* est la conférence internationale, organisée par l'OSGeo, sur les logiciels open source dédiés à la manipulation de données spatiales.

sous-représentation dans l'encyclopédie collective, alors que cette thématique fait l'objet d'études universitaires très avancées. Leur pratique de contribution à un sujet sous-représenté est l'occasion d'analyser les questions de gouvernance dans le cas spécifique de lacunes dans l'encyclopédie. Elles observent que « certaines entrées agissent comme des objets-frontières, des lieux de contestation entre différentes communautés de pratiques » (Gauthier et Sawchuk, 2018). Lors de leurs contributions, elles ont été confrontées à l'enfermement qu'imposent les règles de plus en plus strictes de Wikipédia qui sont lues « à la lettre » par les contributeurs qu'elles qualifient d'experts. Ces contributeurs sont des experts des règles de Wikipédia : « Être expert de Wikipédia a été la clé pour gagner le débat, alors que l'expertise sur le sujet n'a pas été considérée pertinente ». L'enfermement qu'elles dénoncent est le résultat de règles de protection qui visent à éviter que l'encyclopédie ne devienne une plateforme pour la publicité d'entreprises, de groupes ou d'institutions. Dans cette perspective, les contributeurs les plus stricts cherchent à se cantonner aux faits et rigidifient les droits de création de nouvelles entrées dans l'encyclopédie.

Les chercheuses soulignent également la survalorisation de disciplines « perçues comme produisant du savoir 'objectif' ». D'après ces dernières, la tendance est à la sélection de sources biomédicales et à l'élimination de contributions de l'ordre des recherches socio-historiques qui prennent en compte « différentes histoires, genres, sexualités, mœurs culturelles, religion, etc. » (Gauthier et Sawchuk, 2018). En outre, leurs travaux et leurs contributions portant sur le féminisme ont suscité une plus forte propension à la suppression et à la confrontation avec certains contributeurs, jugeant leurs productions biaisées. Les chercheuses ont donc été confrontées à la standardisation du plus grand commun numérique. Elles invitent à réfléchir sur les « stratégies à adopter pour contrer l'encroûtement de ces biais dans la gouvernance de Wikipédia » (Gauthier et Sawchuk, 2018). À travers cette expérience jugée décevante, les chercheuses mettent en évidence la difficulté de rediscuter des règles devenues aujourd'hui garantes de la fiabilité de Wikipédia. Elles ont notamment été confrontées à la règle dites de « notoriété », exigeant que les sources citées soient indépendantes (en termes de conflit d'intérêt) du sujet de l'article. Cette règle garantie la vérifiabilité ainsi que la pérennité du sujet, mais exige également sa diffusion. Si la règle semble adaptée aux visées du projet, en tant qu'encyclopédie, la relation de pouvoir entre experts des règles et néophytes de ces règles semble mettre en danger le commun qu'est

Wikipédia. La difficulté est même accrue avec le durcissement et la complexification des règles et l'entrée de nouveaux contributeurs dans le projet.

Le second exemple qui illustre les conséquences des transformations rapides que connaissent les communs numériques à succès, concerne la gouvernance du projet OSM. Si OpenStreetMap nous semble moins soumis à cette « bureaucratisation », l'augmentation du nombre de contributeurs, inégalement impliqués, impacte directement la gouvernance du système qui devient plus verticale. Nous avons mis en évidence que la faible cohésion du groupe de contributeurs ne s'expliquait pas par les conflits qui ont opposé les membres du groupe. Elle tient plutôt au fait que les interactions restent cantonnées à un petit nombre d'individus qui échangent sur les forums ou les listes de discussions. Cette faible représentativité dans les échanges se retrouve dans la constitution du bureau (*board*) de la Fondation OSM.

Il convient de souligner que la Fondation OSM se définit comme une organisation à but non lucratif dont la visée est la croissance du projet OpenStreetMap. Elle a pour mission d'agir en tant qu'entité légale pour le projet OpenStreetMap et d'être la dépositaire des serveurs informatiques et des services nécessaires pour héberger le projet OpenStreetMap. Elle a également pour objectif de collecter des fonds pour le projet et d'assurer la communication entre ses groupes de travail, les contributeurs et des interlocuteurs extérieurs comme la presse et les entreprises. Fondée en 2007, la Fondation a d'abord eu le rôle de garante de la protection légale du projet et de la pérennité de son système informatique, indépendamment de son premier concepteur Steve Coast. Désormais, la Fondation est un interlocuteur privilégié pour les entreprises qui, comme nous l'avons vu dans le troisième chapitre, construisent parfois leur modèle économique à partir de la base de données OSM. Dans une perspective de valorisation de la ressource, qui croît avec l'intérêt porté par le plus grand nombre, la Fondation intervient dans l'explicitation et l'application de la licence ODbL (au travers de son groupe de travail sur la licence, le *Licensing Working Group*) et dans la résolution de conflits majeurs (au travers de son groupe de travail sur les données, le *Data Working Group*). La Fondation est dirigée par un bureau exécutif dont au moins un tiers des membres doit être renouvelé tous les ans durant des élections accessibles aux membres de la Fondation. Pour devenir membre, une cotisation de quinze livres sterling est exigée.

Compte tenu du rôle croissant occupé par la Fondation pour la pérennité du projet OSM, nous nous sommes intéressés à la constitution du bureau de la Fondation depuis sa création. En effet alors que la Fondation a d'abord été constituée afin de garantir la pérennité du parc informatique hors de la seule personne de Steve Coast, elle est devenue progressivement l'interlocuteur principal des acteurs externes à OSM, eux-mêmes de plus en plus nombreux. Nous procédons à partir des informations publiées sur le *wiki* OSM (*Fondation*, 2018) à un décompte des candidats et des élus ainsi qu'à une classification rapide des professions des élus à partir de leurs déclarations publiques (sur Twitter, le *wiki* OSM et leurs sites internet personnels). De 2007 à 2018, cinquante-huit personnes se sont portées candidates pour intégrer le bureau de la Fondation et vingt-sept ont été élues pour une durée moyenne de trois ans[263]. On perçoit clairement le contraste entre le nombre d'individus engagés dans la gouvernance du projet OSM et le nombre de contributeurs : en mars 2019 près de cinq millions de comptes différents existaient et environ cinq mille comptes ont servi à contribuer durant la seule journée du 19 mars 2019[264]. En onze ans, alors qu'en une seule journée, cinq mille comptes ont pu être actifs, seules cinquante-huit personnes se sont portées candidates au bureau de la Fondation. **Parmi ces vingt-sept élus, onze d'entre eux ont ou avaient, durant leur présence au bureau, des activités professionnelles associées à l'utilisation de la base de données OSM** (cf. Annexe 2). Ces activités se distinguent comme suit :

- Deux élus actuels sont membres de l'association *Humanitarian OpenStreetMap Team* (HOT). Le premier est également employé de l'entreprise Mapbox tandis que le second est employé de la Fondation Wikimédia.
- Les neuf autres élus sont les créateurs ou les employés d'entreprises directement utilisatrices des données OSM comme les entreprises CloudeMade, Geofabrik, Mapzen, Maps.me ou Telenav.

Nous observons que les premiers membres de la Fondation sont des géographes, des géomaticiens et des spécialistes des nouvelles technologies associées à Internet. Le projet n'était donc, dès sa création, pas celui d'amateurs mais bien celui d'experts cherchant à créer un nouveau procédé de création d'informations géographiques. Nous notons également que seulement quatre personnes ne disposent pas de compte sur le *wiki* OSM ou sur Twitter les associant à leurs activités d'élu. Enfin, la faible représentativité des contributeurs (représentativité de genre, de langue maternelle et de

[263] Durée cohérente avec la règle du tiers réélu chaque année pour un bureau composé jusqu'à huit membres.
[264] Données disponibles sur https://osmstats.neis-one.org, consulté en mars 2019. Ce recensement ne compte que les contributions à la base de données et ne prend pas en compte les contributions aux logiciels ou au *wiki* par exemple.

profession) s'accompagne d'une faible participation aux élections. Durant l'élection de décembre 2018, parmi les 1057 électeurs, on a dénombré 680 suffrages exprimés (soit 64%). Ce chiffre faible en termes d'électeurs, compte tenu de l'ampleur démographique du projet, a entraîné la déstabilisation du processus de vote par l'entreprise GlobalLogic[265]. En effet, une hausse importante et inhabituelle du nombre d'inscriptions a été observée durant le mois précédant les élections. Plus de cent personnes se sont inscrites depuis la même adresse IP (adresse des locaux de GlobalLogic en Inde) en quelques heures, le 15 novembre 2018. Les inscriptions ont été signalées immédiatement comme suspectes de par leur simultanéité, la forme automatisée que présentait le renseignement des formulaires d'inscription, l'absence de compte de contributeurs pour une partie et cette adresse IP commune. Finalement, après enquête (Friedl et Rischard, 2018), la Fondation décide de ne pas exclure les membres de l'association mais elle refuse leur participation à l'élection de 2018 car le délai de déclaration était dépassé. Comme le souligne le rapport concluant l'enquête réalisée par la Fondation, celle-ci a été chanceuse d'avoir détecté cette campagne d'inscription d'autant que la procédure d'inscription massive, menée par l'entreprise, a été « bâclée » : l'énorme afflux, de dernière minute, de 100 membres a fortement dépassé le nombre moyen de contributeurs actifs procédant à une demande d'association à la Fondation (Friedl et Rischard, 2018). La Fondation a conclu à une campagne orchestrée et dirigée par GlobalLogic visant à enregistrer massivement leurs employés. Néanmoins, elle ne fait que poser l'hypothèse d'une tentative de déstabilisation des élections sans présumer d'un candidat favori de l'entreprise.

Compte tenu du faible nombre de candidats, il apparaît que les contributeurs à OSM sont très majoritairement non impliqués dans la gouvernance officielle du projet. Ce constat est devenu l'un des sujets les plus abordés dans les professions de foi pour l'élection de 2018 (*Foundation/AGM18/Election to Board*, 2018). Pour cette élection, sept individus se sont porté candidats et six ont rédigé une profession de foi sur le *wiki*. Chacune des professions de foi aborde l'idée qu'un fossé qui sépare la Fondation de la communauté et la nécessité de construire des ponts (« *building a bridge* ») ou de

[265] L'entreprise GlobalLogic a été fondée en 2000 en Californie. Elle présente une large expertise dans le domaine numérique, du design à la conception d'objets connectés. Le rapport qui suit l'enquête réalisée par la Fondation à son propos souligne que « quelques anciens employés de Mapbox ont travaillé à GlobalLogic et sont toujours en contact avec d'anciens collègues. Mapbox est une société américaine très active dans sa contribution à OSM. Elle gérait auparavant une équipe informatique travaillant sur OSM à Bangalore » (Friedl et Rischard, 2018, traduction de l'auteure).

combler le vide (« *filling the gap* ») entre les deux groupes qui ne doivent en faire qu'un. Les candidats défendent l'idée d'une Fondation comme soutien des communautés locales, davantage responsable de la gouvernance à une autre échelle que la seule institution internationale. Dès 2008, une réponse administrative à cette volonté de redistribution locale de l'autorité a été apportée avec la conception des chapitres locaux (*local chapters*).

Ces associations locales sont mandatées par la Fondation et peuvent agir comme les représentants locaux pour les relations avec les gouvernements locaux, les entreprises et les médias. En 2019, six associations locales ont, en accord avec la Fondation, joué ce rôle : OpenStreetMap United Kingdom au Royaume-Uni, Swiss OpenStreetMap en Suisse, Wikimedia Italia en Italie, OpenStreetMap á Íslandi en Islande, l'association FOSSGIS d'Allemagne et OpenStreetMap France. Les professions de foi en vue de l'élection de 2018 faisaient valoir l'intérêt d'une généralisation du procédé de décentralisation. Elles expriment également l'exigence d'un encouragement à la création d'associations partout dans le monde, pour répondre aux besoins spécifiques locaux, tels les besoins matériels, de traduction ou logiciels. Nous relevons l'engagement de Tobias Knerr, élu en 2018, d'intégrer les contributeurs à OSM comme membres de la Fondation. Il s'agit d'inciter les contributeurs à s'impliquer dans la gouvernance du projet notamment par le vote. Joost Schouppe, l'autre élu, se concentre sur une plus grande représentativité des contributeurs, notamment non-anglophones. Tout comme Tobias Knerr, il considère que la communication au sein du groupe mondial est une condition nécessaire à la poursuite du projet. Dans leurs professions de foi, ils déplorent tous les deux la faible diffusion des échanges, qui ont eu lieu majoritairement sur les listes de discussion. Jooste Schouppe milite pour une plus grande diversité de médiums, comme la réalisation d'enquêtes, la rédaction de textes en collaboration et l'utilisation de nouvelles plateformes de discussions structurées (Kialo[266], par exemple). Enfin, il signale qu'une certaine sévérité s'est instaurée dans les discussions en ligne et encourage les « grands noms » (« *big names* ») de la communauté à prendre conscience du poids de leurs mots, leurs interventions maladroites risquant, en effet, d'accentuer la réticence des nouveaux contributeurs à échanger.

[266] https://www.kialo.com/

Ces deux exemples manifestent certaines lacunes dans la construction des communs numériques que sont OSM et Wikipédia. Les deux projets se sont développés très rapidement et leurs ambitions mondiales et pluriculturelles ne sont pas encore atteintes, particulièrement en termes de gouvernance. Elinor Ostrom soulignait déjà, en 1990 et en 2007, l'importance de l'invention de règles originales et spécifiques au commun. Elle soulignait également que plus le commun était grand, plus une organisation avec des composées d'unités imbriquées jusqu'au niveau local, était nécessaire à la pérennité du commun. Nous avons pu constater à travers ces deux exemples que les projets OSM et Wikipédia doivent renouveler leur organisation pour poursuivre le développement de leurs ressources.

En 2019, compte tenu de la faible représentativité des contributeurs et du faible nombre de candidats, dans quelle mesure le projet OSM est-il un commun ?

Alors que nos analyses du projet OSM nous ont mené à qualifier OSM de commun informationnel, le projet présente une organisation détendue. Les contributeurs délaissent la gouvernance du projet à des experts, professionnels du domaine de l'information géographique ou de l'informatique. Comme nous venons de l'exposer, la gouvernance en place est distincte des observations et des recommandations d'Elinor Ostrom. Ce constat peut s'expliquer de deux façons. D'une part, en s'inscrivant sur le site openstreetmap.org, les contributeurs ne s'engagent pas à appartenir à un groupe organisé de manière hiérarchique et décentralisé. D'autre part, la ressource est enrichie par un ensemble d'individus qui s'expriment dans différentes langues - ce qui limite la capacité à participer aux échanges majoritaires en langue anglaise. Nous reprenons les recommandations d'Elinor Ostrom (1990) pour la constitution de communs pérennes afin d'identifier les éventuels risques qui peuvent peser sur l'avenir du projet :

Recommandations (Ostrom, 1990)	**Règles et organisation dans OSM**
1. Des frontières sont clairement définies pour chacun des membres s'agissant de leurs droits sur la ressource.	Dans le cadre d'OSM, la licence ODbL est le texte de référence intégrant le concept de *copyleft* comme cadre légal d'accès à la ressource.
2. L'adéquation est requise entre les règles d'accès et d'utilisation de la ressource et les circonstances propres au commun.	Le projet OSM s'intègre dans le mouvement du Libre. L'accès sans limite aux données et la participation de chacun sont des conditions inhérentes au projet. Le caractère non rival et

	non-exclusif de la base de données permet une exploitation sans limites, autres que le respect de la licence, de la base de données.
3. Des choix collectifs permettent la participation de la majorité des usagers au processus de prise de décision.	Les systèmes de vote pour l'intégration de nouvelles clés existent. Néanmoins, le résultat du vote n'est pas prescriptif. Les usagers peuvent devenir membres de la Fondation et voter pour l'élection des membres du bureau comme pour certaines propositions du bureau.
4. Une évaluation efficace et continue est requise de la part des agents qui font partie du groupe des usagers ou qui sont responsables devant eux.	L'évaluation des comportements des membres porte principalement sur leurs contributions. Elle est rendue possible par l'association d'un contributeur aux modifications de la base auxquelles il procède. L'accès immédiat à l'historique des données simplifie également l'évaluation efficace et continue. Outre le *Data Working Group* de la Fondation, aucun contributeur n'est désigné responsable de l'évaluation. Les contributeurs en désaccord peuvent échanger par le moyen de la messagerie interne à OSM ou par le moyen de la boîte de dialogue disponible pour chaque *changeset*.
5. Des sanctions graduées sont prévues, en fonction de la gravité de l'attaque de la ressource et des connaissances de l'individu.	Comme nous l'avons montré dans le deuxième chapitre, les sanctions sont graduées. Le plus souvent, un contributeur s'adresse à un contributeur à propos d'une édition qu'il juge inadéquate. Lorsque le contributeur n'est pas en mesure d'expliquer les motifs de son écart au modèle ou lorsque son édition est manifestement malveillante, les rares administrateurs (essentiellement membres du *Data Working Group*) suppriment l'édition (*revert*) et coupent les droits d'édition pour un temps donné au contributeur sanctionné.
6. Les mécanismes de résolution des conflits sont peu coûteux et faciles d'accès.	Toute édition peut être corrigée. C'est le premier mécanisme de résolution des conflits. Par ailleurs, le *revert* ne représente aucun coût mais est limité à certains administrateurs.
7. La reconnaissance de ces modes d'organisation est assurée par des structures supérieures, comme les instances gouvernementales.	Dans le deuxième chapitre nous avons exposé l'exploitation croissante des données OSM par certaines institutions et notamment par les villes. Par ailleurs, des acteurs comme Etalab en France entretiennent des relations d'échanges avec les associations locales.
8. Pour les biens communs de grande dimension, une organisation en poupées gigognes est mise en place, avec des unités imbriquées jusqu'au niveau local.	Cette recommandation rejoint l'organisation décentralisée d'une production entre pairs. L'instauration des chapitres locaux mais également des associations locales et des réunions plus informelles peuvent être perçues comme les prémisses de cette organisation en poupées gigognes.

Figure 69 Comparaison des recommandations d'Ostrom et l'organisation d'OSM (Hayat, 2019)

L'analyse de l'organisation autour du commun informationnel OSM permet d'entrevoir différents niveaux d'engagement entre les contributeurs. Cette situation est la conséquence du succès du projet. Selon nous, si le commun conçu par Elinor Ostrom n'est pas complètement transposé, OSM n'est pas, pour autant, un échec. Les recommandations 1, 2, 4, 5, 6 et 7 (de la figure 69) sont atteintes ou en cours d'établissement au sein d'OSM. Au contraire, les recommandations 3 (des choix collectifs permettant la participation de la majorité des usagers au processus de prise de décision) et 8 (pour les biens communs de grande dimension, une organisation en poupées gigognes, avec des unités imbriquées jusqu'au niveau local), ne seront probablement que partiellement suivies. Ces deux recommandations exigent des contributeurs qu'ils s'intègrent à des processus de décision, alors qu'ils sont conçus comme non prescriptifs. La faible implication des contributeurs dans les processus de décision correspond aux différents niveaux d'engagement qui ne peuvent être homogénéisés entre l'ensemble des cinq millions d'utilisateurs enregistrés sur openstreetmap.org.

Le projet OSM consiste à l'enrichissement d'un commun informationnel par plusieurs communautés faiblement centralisées par la Fondation OSM. Le mode de production construit par les contributeurs correspond à la définition de la production par les pairs résumé par Mélanie Dulong de Rosnay dans Orsi et *al.*, 2017. « La production par les pairs est un mode de travail qui s'appuie sur les possibilités techniques de collaboration en ligne […]. Ce processus social et politique d'intelligence collective est une forme de création et de partage en ligne, qui associe trois caractéristiques (Benkler, 2013) : la décentralisation de la conception et de l'exécution de problèmes et de solutions ; une participation fondée sur diverses motivations, notamment non monétaires ; et un mode de gouvernance et de management qui ne s'appuie ni sur le contrat ni sur la propriété » (Mélanie Dulong de Rosnay dans Orsi et *al*., 2017). En ligne comme lors de réunions physiques, les contributeurs échangent essentiellement sur des questions spécifiques liées à un territoire ou à l'usage d'un *tag*. La décentralisation nécessaire, compte tenu de l'ampleur internationale du projet, a entraîné un désintérêt fort de la grande majorité des contributeurs qui délaissent la gouvernance du projet à l'exception d'une très faible

minorité. Les trente élus au bureau de la Fondation en dix ans de projets, constituent une élite coupée de la masse des contributeurs et tendent ainsi vers une organisation plus verticale que celle voulue au départ.

Conclusion du chapitre 4

L'ensemble socio-technique d'OSM que nous avons examiné a été conçu dans les années 1960. Aujourd'hui, il répond à des usages pour lequel il n'était pas nécessairement conçu à l'origine. Les travaux des premiers chercheurs et des premiers développeurs relatifs aux techniques des SIG ont certes permis une amélioration des possibilités logicielles mais également des ergonomies des outils. Néanmoins, des utilisateurs peu sensibles au défi que représentaient, au début des SIG, les outils en ligne de commandes, mais volontaires pour les employer, sont progressivement entrés dans la sphère de la géomatique grâce à ces améliorations. Par ailleurs, les techniques de production sur le Web ont suivi une même trajectoire qui a permis une production plus confortable de données, de plus en plus personnalisables, pour les utilisateurs de services Web. Ces techniques ont également permis une conception plus rapide et plus fonctionnelle de ces services. Ces perfectionnements techniques reposent, selon nous, sur les relations sociales entre chercheurs et entre internautes. **Les réponses techniques comme le mail, IRC, Usenet, à la problématique d'échange des nouveaux supports numériques de documents, ont permis de nouvelles inventions par des groupes partageant des objectifs communs**.

OpenStreetMap est un projet pour lequel des contributeurs ont construit collectivement, en ligne, une infrastructure technique originale. Néanmoins l'organisation sociale n'est pas à la hauteur de l'envergure mondiale du projet s'il se donne pour objectif le fonctionnement en collectif. Les contributeurs à OSM ont emprunté au Libre l'ouverture totale de ses ressources. Ils ont emprunté aux communs le principe d'une organisation réglementée, structurée et décentralisé. Cet assemblage n'est pas achevé. Il s'agit pour les acteurs forts d'OSM, parmi lesquels les membres de la Fondation, d'ouvrir un débat approfondi sur la structure à venir du projet. **S'agit-il d'un projet ouvert et dirigé par un ou quelques individus charismatiques, ou s'agit-il d'une organisation communautaire gouvernée par l'ensemble des membres ?** Il revient à l'intelligence collective d'affronter ces questions et de produire des propositions novatrices et propulsives.

PARTIE II. Conclusion

Dans la première partie de ce travail, nous nous sommes attachées à identifier les éléments explicatifs de la construction de l'écosystème socio-technique original qu'est OSM. Outre des originalités réelles, nous avons souligné que le système de production ainsi construit s'appuie sur ensemble de procédés de production scientifiques traditionnels. Ces premières conclusions sur le projet OSM nous apportent l'éclairage nécessaire à la réalisation d'une cartographie professionnelle sur support numérique et papier que nous détaillons dans le premier chapitre de cette seconde partie (chapitre 3). Nous avons ainsi commencé l'exploration de l'exploitation industrielle de la base de données OSM, objet de cette seconde partie, par l'exposition des procédés de création cartographiques conçus chez Michelin.

Le travail chez Michelin a consisté à mêler les conclusions obtenues au terme de la première partie de ce travail et la réalisation exploratoire de projets cartographiques ; ces explorations ont été détaillées dans le troisième chapitre Ces expérimentations ont pu mettre en avant la nécessité d'avoir un processus de création fortement connecté à l'écosystème du projet OSM pour permettre une exploitation experte de sa base de données. Poursuivant la recherche, il a été possible d'inventorier les avantages et les inconvénients d'un tel changement de méthodes. Dans un processus de création à caractère éditoriale les données libres sont les plus adaptées. En effet, dans le cas de Michelin, la possibilité de se placer en producteur des données et d'en proposer une représentation cartographique sur support papier correspond à un retour aux pratiques de cartographies qui ont fait le succès de l'entreprise. La source libre OSM représente un attrait particulier pour une utilisation à but touristique tant ses données peuvent être adaptées aux spécificités locales. En outre, le cartographe qui aura préparé sa documentation est susceptible, s'il dispose des connaissances nécessaires, d'enrichir la base de données OSM, en cas d'absence de données nécessaires à sa carte. Dans le processus de création de cartes numériques, c'est l'actualité des données qui prime. Au contraire des données libres, toujours accessibles, constamment mises à jour par les contributeurs, les délais de livraison des données propriétaires impliquent une période de plusieurs semaines durant laquelle l'information source est figée dans le temps. Cependant, l'utilisateur de données libres s'expose au vandalisme puisqu'aucun

engagement contractuel ne lui garantit un jeu de données vérifié. Enfin, d'un point de vue technique, prendre le tournant du Libre présente autant d'avantages que d'inconvénients. D'un côté, l'entreprise choisit des outils conçus pour être interopérables et entretenus par des développeurs soucieux de produire des outils qui suivent les évolutions de l'écosystème dans lequel ils s'implantent. De l'autre, employer des outils ouverts exige davantage de ressources en temps et en moyen internes à l'entreprise et celle-ci renonce aux services après-vente du logiciel, argument commercial principal du système propriétaire. Ainsi, nous avons exposé que le changement de source de données implique un changement fort de méthodologie au sein d'une entreprise et n'est pas sans contrainte. Les avantages éditoriaux et méthodologiques consécutifs à l'intégration dans l'écosystème OSM nous semble cohérents à une production cartographique thématique. Enfin, il nous semble qu'aucune économie n'accompagne un tel changement car l'économie de coût que représente l'utilisation de données libres est transférée sur l'investissement en personnels.

Après cette analyse du bouleversement que représente pour une entreprise traditionnelle le tournant vers une source libre, nous avons exploré les conséquences de l'entrée des acteurs économiques dans l'écosystème OSM. L'écosystème d'OSM en est-il déstabilisé ? La réponse est partielle. En effet, le projet n'a finalement jamais atteint de stabilité tant il croît depuis sa création. A la recherche d'informations de contextualisation, nous sommes remontés jusqu'aux années 1960, début de l'informatique telle que nous la pratiquons. Nous avons exposé des éléments explicatifs de la constitution du Géoweb, l'environnement socio-technique du projet OSM. Deux éléments fondamentaux apparaissent à la lecture de notre frise chronologique : d'une part, les dispositifs techniques numériques d'échange sont des espaces de développement d'autres techniques numériques d'échanges et d'autre part, les acteurs de ces dispositifs plaident pour leur standardisation. Ce sont ces deux processus socio-technique qui ont permis la création d'une base de données éditable depuis tout terminal connecté au réseau Internet dans le monde.

En même temps que ces techniques d'échange se développaient, des communautés d'abord de scientifiques, puis davantage diversifiées professionnellement, se constituaient autour de ces dispositifs. Ces communautés ont été les premières à promouvoir les principes de libre accès et de libre diffusion des connaissances qu'elles produisaient. Ceux-ci sont mis en application par le moyen de règles légales de partage

(les licences) et par la progressive organisation des communautés selon des formes plus ou moins hiérarchisées. La communauté d'OSM bien que plus tardive reprend ces principes. OSM choisit comme organisation une forme horizontale menant à la libre modélisation et à la libre contribution à la ressource entretenue par la communauté. Cependant, en même temps que le projet croît, une hiérarchie se distingue, nécessaire aux interactions avec des acteurs extérieurs au projet. OSM prend alors une forme qui répond à son expansion, quelques personnalités portent la voix du groupe. En outre, comme nous l'avons au terme du quatrième chapitre, les membres du groupe ne donnent que rarement leurs voix délibérément à ces personnalités.

L'organisation du projet OSM n'a pas atteint sa stabilité, néanmoins, les discussions se poursuivent en ce sens et nous présumons que les centaines de milliers de contributeurs actifs trouveront une forme d'organisation suffisamment souple pour répondre à la fois aux risques des conflits d'intérêts et à l'absence d'engagement conscient de participer dans la gestion de la communauté. Ces contributeurs peuvent en effet être participants en tant que producteurs (de données ou de logiciels) mais non désireux d'administrer l'ensemble du système.

CONCLUSION GÉNÉRALE

Les résultats que nous avons présentés tout au long de ce travail de thèse sont l'aboutissement d'une recherche menée selon différents points de vue. D'abord, celui d'une contributrice à la base de données OpenStreetMap, ensuite celui d'une cadre de l'entreprise Michelin, responsable de projets de réalisations cartographiques à partir de la source de données libres OSM et enfin, celui d'une chercheuse s'interrogeant sur les perspectives opérationnelles et socio-économiques d'un commun tel que peut l'être OSM.

À l'image de l'organisation de la thèse, nous souhaitons conclure ce travail en suivant quatre parties. Nous commençons par qualifier le projet de commun informationnel et poursuivons en reprenant nos observations selon lesquelles la communauté a construit un système encore instable au profit du commun. Ensuite, nous rappelons l'importance du courant du Libre dans la constitution de la communauté d'OSM. Les analyses que nous avons menées sur l'écosystème socio-technique d'OSM ont nécessité une mise en contexte et ainsi une exploration de la naissance et de la diffusion du courant de pensée du Libre. La communauté OSM créatrice d'un projet original de base de données géographique libre réalise les principes du Libre et des communs en s'en détachant parfois. Enfin, nous revenons sur notre question initiale de recherche consistait à évaluer l'utilisabilité de cette base de données géographique libre pour de la cartographie industrielle. Nous mettons en valeur les conséquences d'un tel changement de source de données pour un acteur traditionnel du domaine de la cartographie.

OSM comme objet à définir

Un commun informationnel, distingué par la nature de sa ressource, possède ses spécificités. Le projet OpenStreetMap vise à la construction et à la pérennisation d'un bien commun informationnel (Ostrom, 1990 ; Coriat, 2015 et Mangolte dans Orsi et *al.*, 2017). Les connaissances contenues dans la base de données OSM, dans l'Encyclopédie Wikipédia ou encore dans le logiciel Qgis, sont non-rivales, non-exclusives et leurs usages par le plus grand nombre les enrichissent, soit grâce à la clause « partage à

l'identique » des licences *copyleft*, soit par le retour des utilisateurs parfois contributeurs. En effet**, la ressource d'un bien commun est destinée à être partagée mais non sans règle**. Ces règles, discutées par les contributeurs qui le souhaitent sur le *wiki* d'OSM, précisent, dans le cas d'un commun informationnel, les conditions de l'enrichissement et de la pérennisation de la ressource en définissant les rôles pour les usagers du commun et les rôles pour les personnes ou entités en charge de son administration. Dans le cas d'OpenStreetMap, ces règles prennent différentes formes, des règles non prescriptives et des règles prescriptives. Les premières consistent en un ensemble de recommandations de *bonnes pratiques* à suivre afin de participer favorablement à la ressource. Ces règles non prescriptives sont notamment conçues pour les relations internes au commun visant l'organisation de la pratique de la folksonomie, à savoir la pratique de classification de ressources en ligne par les internautes à l'aide d'étiquettes ou *tags*. Les règles prescriptives quant à elle, concernent la licence ODbL et certaines décisions de la Fondation. Celle-ci est en mesure d'exclure des contributeurs (lorsqu'ils font acte de vandalisme), de porter l'intégrité d'OSM devant les tribunaux en cas de fraude de la licence et de publier des directives à destination d'acteurs extérieurs (comme l'*Organised Editing Guidelines*). Ces directives sont d'autant plus fondamentales que l'on assiste à l'augmentation forte de la participation des utilisateurs institutionnels ou industriels aux publications d'envergure nationale (Michelin, les mairies telle que celle d'Ottawa et les instituts nationaux de cartographie) et parfois mondiale (Mapbox, Maps.me, Mapillary, TeleNav).

OSM comme système mouvant : discontinuités entre les acteurs extérieurs et les objectifs initiaux du projet OSM

Nous avons longuement exposé le hiatus de plus en plus important qui se forme entre les acteurs économiques et les contributeurs particuliers. Alors que le projet OSM se donne, pour seule limite, le respect de la vie privée et la description d'informations pérennes ne relevant jamais d'un jugement de valeur (ce restaurant est de qualité, par exemple), les acteurs économiques choisissent d'exploiter OSM afin de répondre à leurs propres intérêts.

Ce décalage se constate d'une part, dans la pratique de contribution et d'autre part, lors de l'engagement dans l'organisation du commun. En effet, les contributions des contributeurs individuels sont diverses, parfois centrées sur une thématique et d'autres fois sur un territoire, elles ne sont dirigées par aucune autre finalité que l'enrichissement du commun. Un contributeur individuel souhaitera, par exemple, mettre en valeur son commerce, réaliser une carte de sa commune ou simplement corriger des erreurs qu'il rencontre à l'usage de services basés sur OSM. Plus ses contributions seront conformes aux pratiques majoritaires plus ses modifications seront valorisées.

La base de données OSM, peut être davantage envisagée par des acteurs économiques comme une ressource à adapter à leurs propres besoins. Nous avons pu constater que cette pratique, par exemple de la part de l'entreprise Mapbox tend, au travers de ses outils, à standardiser selon ses propres conceptions le renseignement des carrefours routiers. Néanmoins, cette pratique n'est pas systématique et s'observe lorsque l'acteur économique vise un type d'objet en particulier. Chez Michelin, par exemple, en nous conformant aux pratiques locales de contribution guadeloupéennes, nous bénéficions des contributions passées.

Ce décalage n'est pas pour autant néfaste pour le projet. Nous concluons qu'il existe une certaine complémentarité entre ces deux pratiques et qu'elle est sans aucun doute nécessaire. En outre, cette complémentarité peut se trouver mise en relation à travers le système de règles et de normes qui définissent les droits et les obligations des membres d'une communauté. Cet ensemble réglementaire, mis en application par chacun ou par un sous-groupe de la communauté, constitue alors le fondement de l'ordre au sein du commun, autrement dit sa gouvernance. Cette gouvernance formelle se réalise par des votes pour l'écriture des règles non prescriptives ainsi que par la constitution progressive de l'autorité de la Fondation. Celle-ci, d'abord, personne juridique utilisée afin de collecter des fonds et de s'assurer du paiement des serveurs nécessaires au projet, s'est progressivement incarnée par des personnalités comme Steve Coast, Kate Chapman, Frederik Ramm et Mikel Maron, membres du bureau de la Fondation. Ces quatre personnalités, membres du bureau de la Fondation OSM, sont, comme près de la moitié des élus du bureau, actifs professionnellement dans des domaines mobilisant directement OSM ou des communs informationnels. Nous trouvons là le deuxième décalage que nous évoquions, les contributeurs individuels sont proportionnellement moins investis que certains

contributeurs représentant publiquement leurs employeurs, dans la définition des règles et des normes qui régissent la mise en relation des usagers et des membres du commun (rôle progressivement laissé à la Fondation).

Face à ces décalages, nous envisageons deux perspectives possibles au processus de définition des règles et des normes. Premièrement, nous considérons que les appels vers une diversification des moyens de sociabilité par des membres élus au bureau peuvent mener à une diversification des participants et à une plus grande expression des points de vue. Deuxièmement, nous formulons l'hypothèse que le projet peut tendre vers une hiérarchisation accrue qui trancherait avec les principes initiaux. Cette seconde perspective tendrait vers une uniformisation des intérêts des membres, pour l'enrichissement de la base de données sous une forme davantage standardisée et pourrait avoir pour conséquence l'abandon de la pratique de contribution par certains membres.

Dans ce travail, nous avons posé l'hypothèse que l'action contributive d'entreprises et d'institutions publiques spécialisées, dans le domaine de l'information géographique, trouble le collectif déjà existant. Nous avons confirmé cette hypothèse dans les troisième et quatrième chapitres. Si l'écosystème d'OSM n'a pas atteint d'équilibre, l'entrée croissante d'acteurs extérieurs, dont parfois la création a précédé l'existence d'OSM et d'Internet, confirme son potentiel en tant que nouvelle source d'informations géographiques. L'intérêt porté par les acteurs extérieurs s'explique d'une part par son originalité en contenu et d'autre part par son système de production qui relève du Libre.

OSM comme cas d'étude du courant Libre

Ce système de production, loin d'être fermé, est à la fois technique et vecteur d'échanges sociaux permettant la production entre pairs. Ancré dans le monde d'Internet le courant du Libre, dont OSM s'inspire largement, exploite les techniques numériques d'échanges afin de partager des relations sociales entre pairs et ainsi s'accorder sur des objectifs d'enrichissement de connaissances. Le courant du Libre peut évoquer *a priori* des principes paradoxaux, d'une part en prônant l'interopérabilité et la normalisation des formats de données et des procédures de création afin de réaliser le partage des connaissances ; d'autre part, des contextes comme OSM, emprunts du Libre donc, se distinguent des processus de création traditionnels et ne s'imposent

aucune règle stricte afin de collecter une plus grande diversité d'informations, reconnaissant ainsi que celles-ci peuvent être modélisées de plusieurs façons. Ainsi, non seulement le projet OSM ne respecte pas les règles habituelles, mais il en a créé de nouvelles dans le but de rendre la base de données éditable par tous. **De ce paradoxe et du potentiel désordre apparent sont finalement issues des données utilisables pour la cartographie.** En effet, alors que les contributeurs d'OSM, mobilisant les techniques du Géoweb se sont montrés inventifs, afin de s'affranchir des autorités institutionnelles et de publier leurs propres productions ils ont conservé des pratiques méthodologiques leur permettant de produire des données de plus en plus fiables. En outre, puisqu'une source de connaissances n'est pas seulement fiable par ses résultats et ses acquisitions mais également par sa capacité à progresser, en corrigeant ses erreurs, nous pouvons affirmer que le projet OSM tend à se fiabiliser. Néanmoins, s'il nous semble que cette tendance est la plus probable (et nous rejoignons ici Mocnik, Zipf et Raifer, 2017), une seconde tendance est à envisager. Nous avons décrit une certaine progression de l'accessibilité technique des outils de contribution (avec le développement des logiciels en ligne iD et MapRoulette ou de l'application Maps.me) permettant la contribution d'un plus grand public. Néanmoins, la simplification de la contribution peut atteindre, à terme, la récupération de contributions involontaires. Nous imaginons ainsi le développement d'applications ludiques ou l'usage de robot conversationnel (ou *chatbot*) pour la collecte de données par la captation de textes en langage naturel. Ces applications transformeraient alors les motivations initiales des contributeurs, c'est-à-dire, participer à un ensemble informationnel, que ce soit pour un usage industriel ou pour le divertissement.

Ainsi, les contributeurs ont construit un ensemble socio-technique ouvert favorisant l'itération au sein d'une communauté qui communique régulièrement. **L'information géographique libre est une production formalisée dans un cadre non institutionnel, dans un écosystème particulier, selon les méthodes éprouvées du travail scientifique, que sont le processus de rectification, la reproductibilité des résultats et la vérifiabilité des informations.**

OSM comme source de données pour la cartographie

Au terme de cette thèse, nous sommes en mesure de conclure qu'il est possible de confectionner des cartes professionnelles, à partir des informations géographiques libres produites par la communauté OpenStreetMap, qui respectent le lien de confiance de longue date établi entre les lecteurs et les producteurs de cartes. La réalisation de cartes professionnelles avec OSM requiert cependant trois conditions. La première est la compréhension du système de production qui permet la constitution et l'enrichissement de la base de données. La deuxième condition requiert que la chaîne de production cartographique, afin qu'elle soit durable, soit constituée d'éléments du même environnement technique que celui de la base de données OSM. Enfin, il est impératif de construire un processus de création qui ne peut être la simple transposition d'une ancienne chaîne. En effet, la chaîne de production cartographique issue d'OSM, conçue chez Michelin, est une synthèse technique qui intègre un tout nouveau processus de production d'information qui exige davantage de réalisations de prototypes et requiert le partage des codes sources et des informations géographiques. Les prototypes sont notamment nécessaires à la réalisation de cartes papiers. Le travail centré sur une extraction géographique réduite de la base de données mondiale, conduit le cartographe à se confronter à l'hétérogénéité spatiale des données et à la variation des interprétations du modèle. Les procédés de préparation automatique des cartes doivent être adaptés à chaque pratique de contribution locale afin de bénéficier des spécificités renseignées par les contributeurs.

La chaîne de production se transforme donc en un nouveau cycle de production assurant la mise à jour des documents cartographiques et un contrôle sans équivalent du cartographe sur les données qu'il représente. Le cartographe professionnel devient un contributeur du projet libre OSM dans le cadre de ses fonctions professionnelles. Nous pouvons ainsi affirmer que l'intégration d'informations géographiques libres à un système de production le transforme radicalement au point qu'il devient un nouveau système de production[267]. Néanmoins, cette transformation peut (et sans doute doit), pour un acteur traditionnel, être progressive. L'interopérabilité entre outils libres, nécessaires à l'utilisation des données OSM et outils traditionnels

[267] Le contributeur professionnel enrichie la base de données avec l'intention d'en faire un produit cartographique. Le professionnel s'interroge alors sur les intentions du contributeur individuel qui n'a pas nécessairement la volonté que les informations qu'il a produite soit formalisée par une carte. (Noucher dans Severo et Romele, 2015)

permet la transition progressive vers une chaîne de production entièrement libre, si le producteur souhaite entièrement intégrer sa démarche dans le Libre.

Produire des cartes papiers et des plans de guides, ainsi qu'un prototype de carte en ligne au sein de l'entreprise Michelin Travel Partner, a été l'occasion de constater des motivations que connaissent certaines entreprises à se tourner vers les informations géographiques volontaires. Michelin Travel Partner partage avec d'autres entreprises comme Mappy, un intérêt à disposer d'informations les plus à jour possible, d'informations nouvelles mais également à supprimer le coût d'achat de bases de données afin de disposer de moyens d'investissement pour le développement de techniques d'exploitation de ces données qui présentent un intérêt spécifique. Enfin, comme pour toute production cartographique, le lien avec l'utilisateur acheteur repose sur la confiance envers l'information diffusée et sur une affection envers la politique de l'entreprise. Michelin Travel Partner a parmi ses missions la diffusion du sentiment de confiance, de sécurité et de sympathie envers les acheteurs des produits du groupe Michelin.

Le travail de recherche que nous concluons ici expose les interrelations qui existent entre les techniques du numérique, les acteurs économiques, les acteurs de la recherche et les internautes. Ces interrelations constituent le système à partir duquel la base de données OpenStreetMap est aujourd'hui produite. En 2004, au lancement du projet, le système présentait une forme moins réticulaire et peu d'acteurs traditionnels de l'information géographique y étaient connectés. À la question de l'utilisabilité de la base de données OSM, nous répondons comme la plupart des chercheurs, qu'elle est réelle en fonction de besoins précis et à condition de s'adapter aux techniques qui lui sont propres. Nous ajoutons que le projet OSM est la démonstration de la faisabilité d'un projet contributif international entre anonymes. Cependant, pour qu'il perdure, le système social qui le soutient doit clarifier les règles de travail entre des acteurs aux intérêts différents. La base de données OpenStreetMap peut être la source unificatrice d'un ensemble de connaissances géographiques institutionnelles et citoyennes et peut s'associer aux projets contributifs de la Fondation Wikimedia à condition de ne pas viser la construction d'une base de données unique et omni-usage. En outre si la base

de données peut devenir un lieu d'expression commun c'est notamment grâce au renouvellement méthodologique qui l'accompagne. En se construisant comme un contexte de production original, OSM est propice à la créativité pour une expression populaire. Lorsque l'on aborde des perspectives d'ouverture vers un très grand public il est nécessaire de s'interroger sur l'utilisabilité des données publiques ouvertes. Celles-ci restent difficiles à exploiter pour des individus qui n'auraient suivi aucune formation relative à la gestion de données. L'accessibilité aux données et aux outils, même libres, n'est pas encore atteinte.

Annexes

Annexe 1 : Les contributeurs aux photos Flickr portant le *tag* Street Art, contribuées 23 mars 2018 au 23 mars 2019 à Paris

Nous nous appuyons sur les photographies de 78 contributeurs recensés ainsi :

user_id : user_name

user_name : link; 17643163@N00 : gwolf; 72746018@N00 : dalbera; 41538865@N06 : marsupilami92; 66944824@N05 : Ithmus; 88985500@N00 : GeekShadow; 27892629@N04 : K_rho; 46721940@N00 : Dmitry Dzhus; 7313591@N02 : Renaud Camus; 48518396@N08 : Oliv' C; 91111169@N00 : Pierre_Bn; 10345599@N03 : Francisco Anzola; 92748110@N00 : GillyBerlin; 85264217@N04 : Retis; 68089229@N06 : oatsy40; 50540289@N03 : Bruno_53; 132002900@N07 : Stephen Ben; 98251082@N00 : andre.m(eye)r.vitali; 103560756@N06 : raoulpictor; 105105658@N03 : Rob Oo; 59487880@N00 : aiva.; 96925387@N00 : Jeanne Menjoulet; 79148003@N00 : NathalieSt; 37583176@N00 : Ninara; 22956445@N04 : Naval S; 9361468@N05 : quapan; 123447827@N05 : Mazuritz; 48539981@N03 : Felix Dance; 144601963@N04 : nahia.peschard; 95891889@N04 : Claude Attard; 11854150@N08 : ancasta1901; 14012466@N07 : paologmb; 162718568@N03 : Gbus25- Passion MobilitÃ©s; 144275152@N03 : monbalint; 24427994@N07 : dreadpiratepatty; 11493299@N00 : dany13; 125575033@N04 : Jacques-BILLAUDEL; 135000407@N03 : _elide; 132033145@N08 : osm_hydrant; 101829364@N08 : revenudebasefr; 25092924@N08 : sashimomura; 135141530@N04 : R. W. Rynerson; 12557829@N00 : Simone Ramella; 149039733@N04 : thibaultlaurens; 21606462@N02 : itia4u; 86284987@N08 : SirisVisual; 7259030@N07 : dlofink; 32544079@N08 : jwood601; 10512081@N00 : eugene_o; 107938954@N05 : VinceTraveller; 45457437@N00 : David Jones; 91094471@N02 : Nicolas Vollmer; 9887700@N08 : Photogestion; 31810462@N07 : Carlos ZGZ; 22882695@N00 : jwalsh; 96142817@N03 : jjmusgrove; 89786546@N00 : Randalfino; 87922801@N08 : grits2go; 10688882@N00 : byzantiumbooks; 34862635@N03 : WASD42; 77987497@N00 : Nguyen Vu Hung (vuhung); 90073582@N05 : bilderkombinat berlin; 108760423@N05 : Esnault Denis; 11028609@N05 : Dane Van; 24266175@N00 : La case photo de Got; 163712169@N03 : https://football-family.de; 39549476@N08 : sisssou; 7313683@N06 : alain_sylvie_75; 67950097@N07 : LoveMattersMost; 85825630@N00 : boklm; 11234074@N05 : Philippe Vieux-Jeanton; 25366466@N05 : Newclick; 32174721@N04 : Ali Sabbagh; 126180832@N02 : Fab Photos Canine et Sportive; 90250231@N05 : bernawy hugues kossi huo; 95274266@N06 : Julien Maury; 50785745@N06 : annie harada; 54159132@N04 : ArtS 119@N04 : Josedu92;

Annexe 2 : Les candidats et élus au bureau de la Fondation OSM de 2007 à 2018

Nom	période d'élection	Activités professionnelles ou associatives liées à OSM	Faible ou aucune proximité avec OSM dans les activités associatives ou professionnelles	Source
Corey Burger	2007		Géographe, entrepreneur	http://www.coreyburger.ca/about-me/
Nick Black	2007	Co fondateur de cloud made		https://cloudmade.com/about-us/
Ulf Möller	2009			
Emilie Laffray	2010		Spécialiste en gestion données, big data	https://www.linkedin.com/in/emilielaffray
Iván Sánchez Ortega	2010		Géomaticien, contributeur à Wikipédia, refuse d'être administrateur	https://github.com/IvanSanchez ; https://wiki.openstreetmap.org/wiki/User:Ivansanchez ; https://en.wikipedia.org/wiki/User:Ivansanchez
Kathleen L. Danielson	2014	Employée chez Mapzen en 2014		https://de.linkedin.com/in/kathleendanielson
Joost Schouppe	2018		Déclare ne pas avoir de conflit d'intérêt dans l'exercice de ses fonctions en tant que sociologie spécialiste dans le développement local : « I'm a sociologist specialised in developing geospatial indicators, aimed towards local government employees who want to understand the neighborhoods they are working for ».	https://www.openstreetmap.org/user/joost%20schouppe
Tobias Knerr	2018		Chercheur (research assistant at the University of Passau). Ne mène pas de recherché direct sur OSM mais peut y être confronté: « Although I do not forsee any conflict of interest arising from my employment, though, it's certainly not impossible. For example, my employer is involved in research projects in fields adjacent to geospatial	https://wiki.openstreetmap.org/wiki/Foundation/AGM18/Election_to_Board/Answers_and_manifestos#Conflicts_of_interest

			data, such as self-driving cars. I believe that a conflict of interest disqualifies a board member from using their voting rights and other privileges in the board's decision-making process, and will therefore make sure to evaluate any issues brought before the board for the potential of a CoI ».	
Richard Fairhurst	2007 et 2011-2012		Cartographe et journaliste	https://wiki.openstreetmap.org/wiki/User:Richard
George James	2007-2008		Insiste sur son autonomie professionnelle vis-à-vis d'OSM « My business has no commercial connections or interests in GIS or cartography or any related field ».	https://wiki.openstreetmap.org/wiki/User:80n
Andy Robinson	2007-2009			
Michael Collinson	2007-2009		Entrepreneur dans les nouvelles technologies	https://wiki.openstreetmap.org/wiki/User:Ewmjc/About_Me
Mikel Maron	2007-2011 et 2015-2018	Employé chez Mapbox et co-fondateur de Humanitarian OpenStreetMap Team		https://www.hotosm.org/people/mikel-maron/
Steve Coast	2007-2013	Co-fondateur de l'entreprise CloudeMade		https://blog.openstreetmap.org/2010/10/05/osm-founder-steve-coast-leaves-cloudmade/
Henk Hoff	2008-2014		Géomaticien	https://ke.linkedin.com/in/henkhoff
Simone Cortesi	2009-2010			
Oliver Kühn	2010-2014	Membre de l'entreprise Skobler, calcul d'itinéraire et cartographie avec OSM, propriété de Telenav, date d'entrée dans l'entreprise inconnue		http://blog.skobbler.com/about-2/
Matt Amos	2011-2013	Employé de Mapzen , pas de date précise		https://uk.linkedin.com/in/matt-amos-94175ba0
Dermot	2011-2014	à l'époque fondateur d'une		https://wiki.openstreetmap.org/wi

McNally		startup basée sur OSM		ki/User:Mackerski
Simon Poole	2012-2013		Employé d'un fournisseur d'accès.	https://sosm.ch/fr/about-2/le-conseil-dadministration/simon-poole/
Frederik Ramm	2012-2018	Fondateur et directeur de Geofabrik		https://www.geofabrik.de/geofabrik/geofabrik.html
Kate Chapman	2013-2018	Co-fondateur d'Humanitarian OpenStreetMap Team (HOT), employée Wikimedia Foundation		https://www.hotosm.org/people/kate-chapman/
Paul Norman	2014-2018		Auto-entrepreneurs en développement logiciel	https://wiki.osmfoundation.org/wiki/Board_Member_Bios
Ilya Zverev	2015-2016	Employé chez Maps.me		https://mapsconf.com/en/cth_speaker/ilya-zverev-2/
Peter Barth	2015-2017			
Martijn van Exel	2015-2017	Employé chez Telenav		https://twitter.com/mvexel
Heather Leson	2017-2018		Ancienne member d'humanitarian OpenStreetMap Team (2013-2017)	https://wiki.osmfoundation.org/wiki/Board_Member_Bios

Figure 70 Candidats élus et leurs activités professionnelles et associatives, en gris quatre élus pour lesquels nous n'avons pu réunir d'informations

Les candidats non élu :

Nuno Caldeira, Geoffrey Kateregga, David Dean, Jo Walsh, Miriam Gonzales, Ryan Peterson, Joseph Reeves, Douglas Ssebaggala, Guido Stein, Guillaume Rischard, Darafei Praliaskouski, Gregory Marler, Randy Meech, Ethan Nelson, Yantisa Akhadi, Wille Marcel, Gonzalo Perez, Serge Wroclawski, Alex Barth, Roland Olbricht, Marek Strassenburg-Kleciak, Kai Krueger, Hurricane Coast, Lars Franke, Thea Aldrich, Derick Rethans, Niccolo Rigacci, Eugene Usvitsky, Grant Slater, Richard Weait et Peter Miller.

BIBLIOGRAPHIE

Livres, articles et rapports

Amelot, X., Couderchet, L., & Noucher, M. (2015). Données institutionnelles et données contributives sur la biodiversité, quelle légitimité? *M@ppemonde, (120)*. Disponible sur : http://mappemonde.mgm.fr/120as4/

Bachelard, Gaston. *Essai sur la connaissance approchée*. 6ème éd. Paris, Vrin, 1986, 310 p.

Bachimont, B. (2000). Engagement sémantique et engagement ontologique: conception et réalisation d'ontologies en ingénierie des connaissances. *Ingénierie des connaissances: évolutions récentes et nouveaux défis*, 305-323.

Bakis, H., & Valentin, J. (2010). Amateurisme cartographique et géographique à l'heure du web 2.0. Questionnement autour de la néogéographie. *Netcom. Réseaux, communication et territoires, (24-1/2),* 109-132.

Balley, S. (2007). Pourquoi et comment décrire la structure d'un jeu de données géographiques. *Le Monde des cartes (194)*, 9-16.

Barron, C., Neis, P., & Zipf, A. (2014). A comprehensive framework for intrinsic OpenStreetMap quality analysis. *Transactions in GIS, 18(6)*, 877-895.

Bernard, C., Le Rubrus, B., Villanova-Oliver, M., Gensel, J., Ysebaert, R., & Salmon, I. (2014). STedi: une infrastructure logicielle pour renforcer la qualité des données territoriales statistiques. *In SAGEO Spatial Analysis and GEOmatics*.

Beuscart, J. S . (2002). Les usagers de Napster, entre communauté et clientèle Construction et régulation d'un collectif sociotechnique. *Sociologie du travail, 44(4)*, 461-480.

Board, C. (1971). Maps as models. *In*: Chorley, J., HAGGET, P. *Models in Geography*, 671-725.

Boeters, R., Arroyo Ohori, K., Biljecki, F., & Zlatanova, S. (2015). Automatically enhancing CityGML LOD2 models with a corresponding indoor geometry. *International Journal of Geographical Information Science, 29(12)*, 2248-2268.

Bourdieu, P. (1975). La spécificité du champ scientifique et les conditions sociales du progrès de la raison. *Sociologie et sociétés, 7(1)*, 91-118.

Brimicombe, A. (2009). *GIS, environmental modeling and engineering*. CRC Press, 320 p.

Brown, M., Sharples, S., Harding, J., Parker, C. J., Bearman, N., Maguire, M., Forrest, D., Haklay, & M., Jackson, M. (2013). Usability of geographic information: current challenges and future directions. *Applied Ergonomics, 44(6)*, 855-865.

Bruhat, T. (1984). Genèse. GRETEL: la messagerie interactive-Histoire d'un piratage. *Réseaux. Communication-Technologie-Société, 2(6)*, 37-47. Disponible sur : https://www.persee.fr/doc /reso_0751-7971_1984_num_2_6_1119

Brunet, R., Ferras, R., & Théry, H. (Dir.), (1992). *Les mots de la géographie: dictionnaire critique* (3ème éd) Paris, RECLUS-La Documentation Française, 2003, 520 p.

Bruns, A. (2013). From prosumption to produsage. *Handbook on the digital creative economy,* 67-78.

Bruns, Axel (2006) Towards Produsage: Futures for User-Led Content Production. *In* Sudweeks, Fay, Hrachovec, Herbert, & Ess, Charles (Dir.) *Cultural Attitudes towards Communication and Technology*, Tartu, Estonia.

Budhathoki, N. R., & Haythornthwaite, C. (2013). Motivation for open collaboration: Crowd and community models and the case of OpenStreetMap. *American Behavioral Scientist, 57(5)*, 548-575.

Cauvin, C., Escobar, F., & Serradj, A. (2007). *Cartographie thématique 2–des transformations incontournables*. Hermès-Lavoisier, 272 p.

Chamberlin, D. D. (2012). Early history of SQL. IEEE Annals of the History of Computing, 34(4), 78-82.

Charlet, J., Bachimont, B., & Troncy, R. (2004). Ontologies pour le Web sémantique. *Revue I3, numéro Hors Série « Web sémantique »*, 43-63.

Chilton, S. (2009, November). Crowdsourcing is radically changing the geodata landscape: case study of OpenStreetMap. *In Proceedings of the UK 24th International Cartography Conference* (pp. 15-21).

Cholley, A. (1936). Éditorial. *L'Information Géographique, 1(1),* 1-4.

Chrisman, N. R. (1991). The error component in spatial data. *Geographical information systems, 1(12)*, 165-174.

Christl, A. (2008). Free software and open source business models. *In Open source approaches in spatial data handling* (pp. 21-48). Springer, Berlin, Heidelberg.

Clément-Fontaine, M. (2016). La régulation de l'Open data. *LEGICOM, 56(1)*, 113-120. doi:10.3917/legi.056.0113.

Coriat, B. (2015). *Le retour des communs: & la crise de l'idéologie propriétaire*. Mayenne, France : Les Liens qui Libèrent, 298 p.

Delestre, N., & Malandain, N. (2017). *Du web des documents au web sémantique*. Éditions KLOG, 208p.

Devillers, R., Stein, A., Bédard, Y., Chrisman, N., Fisher, P., & Shi, W. (2010). Thirty years of research on spatial data quality: achievements, failures, and opportunities. *Transactions in GIS, 14(4)*, 387-400.

Duféal, M., Jonchères, C., & Noucher, M. (2016*). ECCE CARTO - DES ESPACES DE LA CON-TRIBUTION A LA CONTRIBUTION SUR L'ESPACE - Profils, pratiques et valeurs d'engagement des contributeurs d'OpenStreetMap (OSM)* . [Rapport de recherche] UMR 5319.

Estellés-Arolas, E., & González-Ladrón-De-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information science, 38(2)*, 189-200.

Estima, J., & Painho, M. (2013, June). Flickr geotagged and publicly available photos: Preliminary study of its adequacy for helping quality control of Corine land cover. *In*

International Conference on Computational Science and Its Applications (pp. 205-220). Springer, Berlin, Heidelberg.

Estima, J., & Painho, M. (2015). Investigating the potential of OpenStreetMap for land use/land cover production: A case study for continental Portugal. *In OpenStreetMap in GIScience* (pp. 273-293). Springer, Cham.

Fischer, H. T. (1982). *Mapping Information: The Graphic Display of Quantitative Information. Cambridge*, Abt books, 1983, 410p.

Fisher, P. F. (1999). Models of uncertainty in spatial data. . *In* Longley, P., Goodchild M., Maguire D., & Rhind D. (dir.). *Geographical information systems: principles and technical issues* (pp. 191-205), (2ème éd). New York, John Wiley & sons, vol.1.

Flanagin, A. J., & Metzger, M. J. (2008). The credibility of volunteered geographic information. *GeoJournal, 72(3-4),* 137-148.

Friedl, S., Rischard, G. (26 December 2018). *Investigation into the Unusual Signups*. Consulté en mai 2019 sur: https://openstreetmap.lu/MWGGlobalLogicReport20181226.pdf

Frischmann, B. M., Madison, M. J., & Strandburg, K. J. (Dir.). (2014). *Governing knowledge commons*. Oxford University Press, 499p.

Gauthier, M., & Sawchuk, K. (2018). Le partage de quels savoirs? Les articles Wikipédia comme objets-frontières. *tic&société, 12(1),* 131-165.

German, D. M. (2006, May). A study of the contributors of PostgreSQL. *In MSR (6),* pp. 163-164).

Girres, J. F., & Touya, G. (2010). Quality assessment of the French OpenStreetMap dataset. *Transactions in GIS, 14(4)*, 435-459.

Glassey, O. (2006). Wikipédia: une encyclopédie des controverses ouvertes. *In* Leresche, JP, Benninghoff, M., Crettaz von Roten, F., Merz, M. *La fabrique des sciences: des institutions aux pratiques*. Lausanne: Presses polytechniques et universitaires romandes, 237-256.

Glasze, G., & Perkins, C. (2015). Social and political dimensions of the OpenStreetMap project: Towards a critical geographical research agenda. *In OpenStreetMap in GIScience* (pp. 143-166). Springer, Cham.

Goetz, M. (2013). Towards generating highly detailed 3D CityGML models from OpenStreetMap. *International Journal of Geographical Information Science, 27(5),* 845-865.

Goodchild, M. F. (2007). Citizens as sensors: the world of volunteered geography. *GeoJournal, 69(4)*, 211-221.

Goodchild, M. F., & Gopal, S. (Dir.). (1989). *The accuracy of spatial databases*. CRC Press.

Goodchild, M. F., & Li, L. (2012). Assuring the quality of volunteered geographic information. *Spatial statistics (1)*, 110-120.

Gröchenig, S., Brunauer, R., & Rehrl, K. (2014). Digging into the history of VGI data-sets: results from a worldwide study on OpenStreetMap mapping activity. Journal of Location *Based Services, 8(3)*, 198-210.

Gruber T. (2009) Ontology. *In* Liu L., Özsu M.T. (Dir.) *Encyclopedia of Database Systems*. Springer, Boston, MA

Gruber, T. R. (1993). A translation approach to portable ontology specifications. *Knowledge acquisition, 5(2)*, 199-220.

Gruber, T. R. (1995). Toward principles for the design of ontologies used for knowledge sharing?. *International journal of human-computer studies, 43(5-6)*, 907-928.

Guérois, M., Feredj, A., Giraud, T., Lambert, N., Viry, M., & Ysebaert, R. (2016). *Rapport final typologie socio-économique des régions frontalières de l'union européenne (2000-2012)*, RIATE et CGET.

Guptill, S.C. & Morrison, J. L. (Dir.). (1995). *Elements of Spatial Data Quality*, Elsevier, Oxford.

Gutiérrez, R. T. (2018). Understanding the role of digital commons in the web; The making of HTML5. *Telematics and Informatics, 35(5)*, 1438-1449.

Haklay, M. (2010). How good is volunteered geographical information? A comparative study of OpenStreetMap and Ordnance Survey datasets. *Environment and planning B: Planning and design, 37(4)*, 682-703.

Haklay, M. (2013). Citizen science and volunteered geographic information: Overview and typology of participation. *In Crowdsourcing geographic knowledge* (pp. 105-122). Springer, Dordrecht.

Haklay, M., Basiouka, S., Antoniou, V., & Ather, A. (2010). How many volunteers does it take to map an area well? The validity of Linus' law to volunteered geographic information. *The Cartographic Journal, 47(4)*, 315-322.

Hall, A., Thebault-Spieker, J., Sen, S., Hecht, B., & Terveen, L. (2018, August). Exploring the Relationship Between Informal Standards and Contributor Practice in OpenStreetMap. *In Proceedings of the 14th International Symposium on Open Collaboration* (pp.10). ACM.

Hardin, G. (1968). The tragedy of the commons. *Science, 162(3859)*, 1243-1248.

Harley, J. B. (1989). Deconstructing the map. Cartographica: *The international journal for geographic information and geovisualization, 26(2)*, 1-20.

Harley, J. B., Woodward, D., & Lewis, G. M. (Dir.). (1987). *The history of cartography (1)*. Chicago: University of Chicago Press.

Harvey, F. (2013). To volunteer or to contribute locational information? Towards truth in labeling for crowdsourced geographic information. *In Crowdsourcing geographic knowledge* (pp. 31-42). Springer, Dordrecht.

Haythornthwaite, C. (2009). Crowds and communities: Light and heavyweight models of peer production. *In 42nd Hawaii international conference on system sciences* (pp. 1-10). IEEE.

Hentschel, M., & Wagner, B. (2010). Autonomous robot navigation based on openstreetmap geodata. *In 13th International IEEE Conference on Intelligent Transportation Systems* (pp. 1645-1650). IEEE.

Hess, C., & Ostrom, E. (2007). *Understanding knowledge as a commons*. Cambridge, Massachusetts, Londres, Angleterre :The MIT Press.

Hombiat, A. (2017). *OF4OSM : un méta-modèle pour structurer la folksonomie d'OpenStreetMap en une nouvelle ontologie.* Informatique. Université Grenoble Alpes, Français.

Hombiat, A., Villanova-Oliver, M., & Gensel, J. (2015). Un nouveau méta-modèle pour rapprocher la folksonomie et l'ontologie d'OSM. *In Colloque International de Géomatique et d'Analyse Spatiale SAGEO,* 103-119.

Howe, J. (2006). The rise of crowdsourcing. *Wired magazine, 14(6)*, 1-4.

Hunziker, J., & Oderbolz, S. (2012). *Gamified Mobile App für die Verbesserung von OpenStreetMap*. Informatique. HSR Hochschule für Technik Rapperswil, Allemand.

Johnson, B. A., & Iizuka, K. (2016). Integrating OpenStreetMap crowdsourced data and Landsat time-series imagery for rapid land use/land cover (LULC) mapping: Case study of the Laguna de Bay area of the Philippines. *Applied Geography, 67*, 140-149.

Jokar Arsanjani, J., & Bakillah, M. (2015). Understanding the potential relationship between the socio-economic variables and contributions to OpenStreetMap. *International Journal of Digital Earth, 8(11)*, 861-876.

Jokar Arsanjani, J., Helbich, M., Bakillah, M., & Loos, L. (2015). The emergence and evolution of OpenStreetMap: a cellular automata approach. *International Journal of Digital Earth, 8(1),* 76-90.

Joliveau, T. (2011). Le géoweb, un nouveau défi pour les bases de données géographiques. *L'Espace geographique, 40(2)*, 154-163.

Karimipour, F., & Azari, O. (2015). Citizens as expert sensors: One step up on the VGI ladder. *In Progress in Location-Based Services* (pp. 213-222). Springer, Cham.

Keen, A. (2007). *The cult of the amateur : how today's internet is killing our culture*. New York: Doubleday/Currency.

Keßler, C., & De Groot, R. T. A. (2013). Trust as a proxy measure for the quality of volunteered geographic information in the case of OpenStreetMap. *In Geographic information science at the heart of Europe* (pp. 21-37). Springer, Cham.

Lambert, N., & Zanin, C. (2012). OpenStreetMap: collaborer pour faire des cartes. *Mappemonde, 107(3).*

Larroche, V., & Dymytrova, V. (2018). L'apport des SIC aux projets Open Data: pour une meilleure prise en compte des réutilisateurs dans la conception des outils de traitement des data. *Communication Organisation, (2)*, 69-80.

Laval, C. (2016). «Commun» et «communauté»: un essai de clarification sociologique. *SociologieS*.

Le Crosnier, H. (2015). Affrontements politiques au sein de l'écosystème numérique. *EcoRev', (2)*, 61-71.

Le Goix, R. (2018) "Analyse spatiale", in: A. Ciatonni & Y. Veyret (Dir.) *Les fondamentaux de la géographie*, pp. 43-67. Paris: Armand Colin.

Leiner, B. M., Cerf, V. G., Clark, D. D., Kahn, R. E., Kleinrock, L., Lynch, D. C., Postel, J., Roberts, L. G., Wolff, S. (2009). A brief history of the Internet. *ACM SIGCOMM Computer Communication Review, 39(5),* 22-31.

Leon, L. F. A., & Quinn, S. (2018). The value of crowdsourced street-level imagery: examining the shifting property regimes of OpenStreetCam and Mapillary. *GeoJournal*, 1-20.

Li, H., Tao, C., Wu, Z., Chen, J., Gong, J., & Deng M. (2017). RSI-CB: A Large Scale Remote Sensing Image Classification Benchmark via Crowdsource Data.

Limpens, F., Gandon, F., & Buffa, M. (2008). Rapprocher les ontologies et les folksonomies pour la gestion des connaissances partagées: un état de l'art. *In 19ème Journées Francophones d'Ingénierie des Connaissances,* 123-134.

Liu, X., & Long, Y. (2016). Automated identification and characterization of parcels with OpenStreetMap and points of interest. *Environment and Planning B: Planning and Design, 43(2)*, 341-360.

Liu, X., He, J., Yao, Y., Zhang, J., Liang, H., Wang, H., & Hong, Y. (2017). Classifying urban land use by integrating remote sensing and social media data. *International Journal of Geographical Information Science, 31(8)*, 1675-1696.

Loidl M., Fischer F., Traun C. (2011). Transactional Map Symbols -At the crossroads of cartography?, In Car A., Griesebner G., Strobl J. (Dir.), *GI-Forum*, Salzburg, Wichman Verlag, p.51-61.

Longley, P. A., Goodchild, M. F., Maguire, D. J., & Rhind, D. W. (2005). *Geographic information systems and science*. John Wiley & Sons.

Macbeth, S., & Pitt, J. V. (2015). Self-organising management of user-generated data and knowledge. *The Knowledge Engineering Review, 30(3)*, 237-264.

Maguire, D. J., Goodchild, M. F., & Rhind, D. W. (1991). Geographical information systems: principles and applications. Two volumes. *In Geographical information systems: principles and applications. Two volumes.* Longman, Harlow/Wiley, New York.

Mangolte, P. A. (2015). Le logiciel libre comme commun créateur de richesses. In Coriat, B.(Dir.), *Le retour des communs. La crise de l'idéologie propriétaire*, Mayenne, France : Les Liens qui Libèrent, 298 p.

Marcoccia, M. (2001). L'animation d'un espace numérique de discussion: l'exemple des forums usenet. *Document numérique, 5(3)*, 11-26.

Marquet, C. (2016). Des services connectés pour améliorer l'accessibilité des gares?. *Espace populations sociétés*. (2016/2).

Maurel, L. (2012). Du web de documents au web de données : la révolution juridique inachevée de l'Open Data. In Calderan, L., Laurent, P., Lowinger, H., & Millet, J.. *Le document numérique à l'heure du web* (pp.155-171), ADBS.

Maurel, L. (2014). Droit d'auteur et création dans l'environnement numérique. *Mouvements, (3)*, 100-108. Url : https://www.cairn.info/revue-mouvements-2014-3-page-100.htm

Mechouche, A., Abadie, N., Prouteau, E., & Mustière, S. (2011). Ontology-Based Discovering of Geographic Databases Content. *In Advances in Cartography and GIScience(1)* (pp. 311-330). Springer, Berlin, Heidelberg.

Mericskay, B. (2016). La cartographie à l'heure du Géoweb: Retour sur les nouveaux modes de représentation spatiale des données numériques. *Cartes & géomatique, 229*, 37-50.

Mocnik, F. B., Zipf, A., & Raifer, M. (2017). The OpenStreetMap folksonomy and its evolution. *Geo-spatial Information Science, 20(3)*, 219-230.

Monnin, A., & Félix, E. (2009). Essai de comparaison des ontologies informatiques et philosophiques: entre être et artefacts. *In Rochebrune'09: Ontologie et dynamique des systèmes complexes, perspectives interdisciplinaires* (pp. 14) Megève, France.

Mooney, P., & Corcoran, P. (2012). Characteristics of heavily edited objects in OpenStreetMap. *Future Internet, 4(1)*, 285-305.

Mooney, P., & Corcoran, P. (2012, April). How social is OpenStreetMap. *In Proceedings of the 15th association of geographic information laboratories for europe international conference on geographic information science*, Avignon, France (pp. 24-27).

Mooney, P., Corcoran, P., & Ciepluch, B. (2013). The potential for using volunteered geographic information in pervasive health computing applications. *Journal of Ambient Intelligence and Humanized Computing, 4(6)*, 731-745.

Mueller, N., Rojas-Rueda, D., Salmon, M., Martinez, D., Ambros, A., Brand, C., de Nazelle, A., Dons, E., Gaupp-Berghausen, M., Gerike, R., Götschi, T., Iacorossi, F., Panis, L. I., Kahlmeier, S., Raser, E., & Nieuwenhuijsen, M. (2018). Health impact assessment of cycling network expansions in European cities. *Preventive medicine, 109*, 62-70.

Neis, P., & Zipf, A. (2012). Analyzing the contributor activity of a volunteered geographic information project—The case of OpenStreetMap. *ISPRS International Journal of Geo-Information, 1(2)*, 146-165.

Neis, P., Goetz, M., & Zipf, A. (2012). Towards automatic vandalism detection in OpenStreetMap. *ISPRS International Journal of Geo-Information, 1(3)*, 315-332.

Neis, P., Zielstra, D., & Zipf, A. (2012). The street network evolution of crowdsourced maps: OpenStreetMap in Germany 2007–2011. *Future Internet, 4(1)*, 1-21.

Noucher, M. (2015). De la trace à la carte et de la carte à la trace: pour une approche critique des nouvelles sources de fabrique cartographique *In* Severo, M., & Romele, A. (Dir.). *Traces numériques et territoires*. Presses des Mines via OpenEdition, 270p.

Nuckelt, J., Rose, D. M., Jansen, T., & Kürner, T. (2013). On the use of OpenStreetMap data for V2X channel modeling in urban scenarios. *In 2013 7th European Conference on Antennas and Propagation (EuCAP)* (pp. 3984-3988). IEEE.

Olson, M. (1965). Logic of collective action: Public goods and the theory of groups. *Harvard economic studies, 124*. Harvard University Press.

Orsi, F., Rochfeld, J., & Cornu-Volatron, M. (2017). *Dictionnaire des biens communs*. Presses Universitaires de France.

Ostrom, E. (1990). *Governing the Commons: The Evolution of Institutions for Collective Action*. Cambridge : University Press.

Ostrom, E. (2011). Plaidoyer pour la complexité. *Ecologie politique, (1)*, 111-121. Doi : 10.3917/ecopo.041.0111. Url : https://www.cairn.info/revue-ecologie-et-politique1-2011-1.htm-page-111.html

Over, M., Schilling, A., Neubauer, S., & Zipf, A. (2010). Generating web-based 3D City Models from OpenStreetMap: The current *situation In Germany. Computers, Environment and Urban Systems, 34(6)*, 496-507.

Paloque-Berges, C. (2012). La mémoire culturelle d'Internet: le folklore de Usenet. *Le Temps des médias, (1)*, 111-123.

Palsky, G. (2013). Cartographie participative, cartographie indisciplinée. *L'Information géographique, 77(4),* 10-25.

Peuquet, D. J., & Marble, D. F. (1990). Arc/info: an example of a contemporary geographic information system. In Introductory Readings *In Geographic Information Systems* (pp. 112-122). CRC Press.

Pouzin, L. (1973). Presentation and major design aspects of the CYCLADES computer network. *In Proceedings of the third ACM symposium on Data communications and Data networks: Analysis and design* (pp. 80-87). ACM.

Psyché, V., Mendes, O., & Bourdeau, J. (2003). Apport de l'ingénierie ontologique aux environnements de formation à distance. *Sciences et Technologies de l'Information et de la Communication pour l'Éducation et la Formation, 10*, 89-126.

Raymond E. (2001). *The Cathedral & the Bazaar Musings on Linus and Open Source by an accidental revolutionnary*, O'Reilly, 208 p

Reichardt, M. & Robida, F. (2019). Why Standards Matter – The objectives and roadmap of the International Open Geospatial Consortium (OGC). *Annales des Mines - Responsabilité et environnement, 94(2),* 25-29. https://www.cairn.info/revue-responsabilite-et-environnement-2019-2-page-25.htm.

Reid, E. M. (1991). *Electropolis: Communication and community on internet relay chat.* University of Melbourne, Department of History.

Reid, E. M. (1996). Communication and community on Internet Relay Chat: constructing communities. High noon on the electronic frontier: *Conceptual issues in cyberspace,* 397-411.

Rey, A., & Morvan, D. (2005). *Dictionnaire culturel en langue française*. Le Robert.

Robinson, V. B., & Tom, H. (Dir.). (1993). *Towards sql database language extensions for geographic information systems*. DIANE Publishing.

Rouet, P. (1991). *Les données dans les SIG*, Paris.

Rousell, A., Hahmann, S., Bakillah, M., & Mobasheri, A. (2015). Extraction of landmarks from OpenStreetMap for use in navigational instructions. *In Proceedings of the AGILE Conference on Geographic Information Science* (pp. 9-12), Lisbone, Portugal.

Sahut, G. (2015). *Wikipédia, une encyclopédie collaborative en quête de crédibilité: le référencement en questions.* Sciences de l'information et de la communication .Université Toulouse Jean Jaurès, français.

Sanjay R., & Joliveau, T. (2009). NeoGeography: An extension of mainstream geography for everyone made by everyone?. *Location Based Services, 3,* 75-81.

See, L., Mooney, P., Foody, G., Bastin, L., Comber, A., Estima, Fritz, S., kerle, N., Jiang, B., Laakso, M., Liu, H. Y., Milčinski, G., Nikšič, M., Painho, M., Pődör, A., Olteanu-Raimond A. M. & Rutzinger M. (2016). Crowdsourcing, citizen science or volunteered geographic information? The current state of crowdsourced geographic information. *ISPRS International Journal of Geo-Information, 5(5)*, 55.

Severo M. (2014), « La liste du patrimoine mondial : une approche par le système d'information », *In Culture et musées* (41), p. 41-70.

Severo, M. (2017). Suivre le médium numérique: les méthodes numériques en SIC. *Revue française des sciences de l'information et de la communication*, (10).

Simon, N., & Alain, M. (1978). *L'informatisation de la société: rapport à M. le Président de la République*. La documentation française.

Stefanidis, A., Crooks, A., & Radzikowski, J. (2013). Harvesting ambient geospatial information from social media feeds. *GeoJournal, 78(2)*, 319-338.

Stvilia, B., Gasser, L., Twidale, M. B., & Smith, L. C. (2007). A framework for information quality assessment. *Journal of the American society for information science and technology, 58(12)*, 1720-1733.

Surowiecki, J. (2005). *The wisdom of crowds*. New York: Anchor Books.

Tapscott, D., & Williams, A. D. (2006). *Wikinomics: How mass collaboration changes everything.* New York: Portfolio.

Toffler, A., & Alvin, T. (1980). *The third wave*. New York: Bantam book

Tomlinson, R. F. (1968). A geographic information system for regional planning. *In Symposium on Land Evaluation, Commonwealth Scientific and Industrial Research Organization*, 200–210.

Touya, G., Antoniou, V., Olteanu-Raimond, A. M., & Van Damme, M. D. (2017). Assessing crowdsourced POI quality: Combining methods based on reference data, history, and spatial relations. *ISPRS International Journal of Geo-Information*, *6*(3), 80.

Trojette, M. A., & Lombard, R. (2013). *Ouverture des données publiques. Les exceptions au principe de gratuité sont-elles toutes légitimes*. Paris, France : La Documentation Française

Truong, Q., Touya, G., & De Runz, C. (2017). Le vandalisme dans l'information géographique volontaire: apprendre pour mieux détecter ? *In SAGEO 2018*, 61-76.

Tukey, J. W. (1961). Statistical and quantitative methodology. *Trends in Social Science*, 84-136.

Turner, A. (2006). *Introduction to Neogeography.* Sebastopol, CA: O'Reilly Media.

Van Oort, P. A. (2006). *Spatial data quality: from description to application*. Wageningen Universiteit, néerlandais

Vieira, L. (2015). Les écosystèmes numériques: le Big Bang. Aux sources des logiques de l'information en réseau. *In* Agostinelli, S., Koulayan, N. (Dir.) *Les écosystèmes numériques et la démocratisation informationnelle: Intelligence collective, Développement durable, Interculturalité, Transfert de connaissances*. Presses des Mines.

Vienne, F., Douay, N., Le Goix, R., & Severo, M. (2017). Les territoires du réseau social facebook: le cas des pratiques de géoréférencements. *Territoire en mouvement Revue de géographie et aménagement. Territory in movement Journal of geography and planning*, (34).

Wallace, T. R. (2011). A new map sign typology for the GeoWeb. *In International Cartographic Conference.*

Westervelt, J. (2005). GRASS roots. *International Journal of Geoinformatics, 1(1).*

Xifaras, M. (2010). Le copyleft et la théorie de la propriété. *Multitudes, (2)*, 50-64.

Yang, A., Fan, H., & Jing, N. (2016). Amateur or professional: Assessing the expertise of major contributors in OpenStreetMap based on contributing behaviors. *ISPRS International Journal of Geo-Information, 5(2)*.

Zhang, J., & Goodchild, M. F. (2002). *Uncertainty in geographical information*. CRC press.

Zielstra, D., & Zipf, A. (2010, May). A comparative study of proprietary geodata and volunteered geographic information for Germany. *In 13th AGILE international conference on geographic information science.*

Zielstra, D., Hochmair, H. H., & Neis, P. (2013). Assessing the Effect of Data Imports on the Completeness of OpenStreetMap–A United States Case Study. *Transactions in GIS, 17(3)*, 315-334.

Sites Web

Bachimont, B. (2006). Qu'est-ce qu'une ontologie ? Consulté en avril 2019 sur : http://www.technolangue.net/imprimer.php3? id_article=280.

Berteaud P (2013). Pour en finir avec de fausses informations sur l'IGN !. Consulté en mai 2019 sur : https://www.lagazettedescommunes.com/205912/pour-en-finir-avec-de-fausses-informations-sur-lign/

Blanc, S. et Fauvel V. (2013). Le rapport Trojette : les acteurs du secteur saluent un bon point de départ (2/2). Consulté en mai 2019 sur : https://www.lagazettedescommunes.com/205588/le-rapport-trojette-les-acteurs-du-secteur-saluent-un-bon-point-de-depart-22/

Cariou, C. & Gaulon-brain,M. Du minitel à l'internet. Consulté en avril 2019 sur : https://www.inaglobal.fr/telecoms/article/du-minitel-linternet

Le Deuff, O. (2006). Folksonomies. Les usagers indexent le web. Consulté en avril 2019 sur : http://bbf.enssib.fr/consulter/bbf-2006-04-0066-002?em_x=22

Stallman R. (2013). Why Open Source misses the point of Free Software .Consulté en mai 2019 sur : https://www.gnu.org/philosophy/open-source-misses-the-point.en.html

Stallman, R. The GNU manifesto. Consulté en avril 2019 sur : https://www.gnu.org/gnu/manifesto.en.html

Vander Wal, T (2004). *Feed on This*. Consulté en avril 2019 sur : http://vanderwal.net/random/entrysel.php?blog=1562

Vander Wal T. (2005).Folksonomy Definition and Wikipedia. Consulté en janvier 2019 sur http://www.vanderwal.net/random/entrysel.php?blog=1750

Wikis

Aerial imagery. (2018, September 2). OpenStreetMap Wiki, . Retrieved 12:49, mai 14, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Aerial_imagery&oldid=1645567.

API. (2019, January 28). *OpenStreetMap Wiki*, . Retrieved 10:30, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=API&oldid=1782017.

Beginners Guide 1.3. (2019, March 2). *OpenStreetMap Wiki*, . Retrieved 07:57, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Beginners_Guide_1.3&oldid=1814709.

Current events. (2018, March 4). *OpenStreetMap Wiki*, . Retrieved 14:49, mai 21, 2019 from Trojette, M. A., & Lombard, R. (2013). Ouverture des données publiques. Les exceptions au principe de gratuité sont-elles toutes légitimes. Cahier, (2), 13-14.

Dhaka Building Import. (2019, March 28). *OpenStreetMap Wiki*, . Retrieved 08:19, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Dhaka_Building_Import&oldid=1828077.

DigitalGlobe. (2018, January 24). *OpenStreetMap Wiki*, . Retrieved 17:01, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=DigitalGlobe&oldid=1557338.

Editing Standards and Conventions. (2019, February 3). *OpenStreetMap Wiki*, . Retrieved 07:53, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Editing_Standards_and_Conventions&oldid=1784987.

Foundation. (2018, December 29). *OpenStreetMap Wiki*, . Retrieved 09:22, mai 20, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Foundation&oldid=1764924.

Foundation/AGM18/Election to Board. (2018, December 15). *OpenStreetMap Wiki*, . Retrieved 11:54, mai 20, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Foundation/AGM18/Election_to_Board&oldid=1712826.

Foundation/Local Chapters. (2019, February 10). *OpenStreetMap Wiki*, . Retrieved 12:18, mai 20, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Foundation/Local_Chapters&oldid=1788249.

FR:Modifications automatisées. (2019, April 21). *OpenStreetMap Wiki,* . Retrieved 13:39, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=FR:Modifications_automatis%C3%A9es&oldid=1844639.

FR:WikiProject France/WikiProject Base Adresses Nationale Ouverte (BANO). (2019, January 27). *OpenStreetMap Wiki,* . Retrieved 08:31, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=FR:WikiProject_France/WikiProject_Base_Adresses_Nationale_Ouverte_(BANO)&oldid=1781698.

https://wiki.openstreetmap.org/w/index.php?title=Current_events&oldid=1581032.

Import/Catalogue. (2019, May 11). *OpenStreetMap Wiki,* . Retrieved 08:12, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Import/Catalogue&oldid=1851055.

Import/Guidelines. (2018, October 9). *OpenStreetMap Wiki,* . Retrieved 08:13, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Import/Guidelines&oldid=1678324.

Inception. (2019, mai 4). Wikidata. Retrieved 10:30, mai 13, 2019 https://www.wikidata.org/wiki/Property:P571.

Key:heritage. (2019, March 4). *OpenStreetMap Wiki,* . Retrieved 09:35, mai 25, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Key:heritage&oldid=1816939.

Key:tidal. (2019, February 23). *OpenStreetMap Wiki,* . Retrieved 09:29, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Key:tidal&oldid=1804765.

Lacking proper attribution. (2019, April 5). *OpenStreetMap Wiki,* . Retrieved 13:27, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Lacking_proper_attribution&oldid=1831027.

Main Page. (2014, July 10). *OpenStreetMap Wiki,* . Retrieved 14:04, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Main_Page&oldid=1060762.

Map Features. (2019, March 10). *OpenStreetMap Wiki,* . Retrieved 10:28, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Map_Features&oldid=1819914.

MAPS.ME. (2019, January 24). *OpenStreetMap Wiki,* . Retrieved 16:57, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=MAPS.ME&oldid=1780022.

Œuvre libre. (2019, avril 24). *Wikipédia, l'encyclopédie libre*. Page consultée le 13:13, avril 24, 2019 à partir de http://fr.wikipedia.org/w/index.php?title=%C5%92uvre_libre&oldid=158710258.

Openstreetmap Foundation (2018), *Organised editing guidelines* [en ligne]. Disponible sur : https://wiki.osmfoundation.org/w/images/6/62/Organised_Editing_Guidelines.pdf

Organised Editing Guidelines. (2019, January 28). *OpenStreetMap Wiki,* . Retrieved 16:33, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Organised_Editing_Guidelines&oldid=1781987.

Pick your mapping technique. (2018, September 20). *OpenStreetMap Wiki,* . Retrieved 07:56, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Pick_your_mapping_technique&oldid=1659356.

Planet.osm. (2019, May 1). *OpenStreetMap Wiki,* . Retrieved 13:34, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Planet.osm&oldid=1848102.

Proposed features/Tidal Rivers. (2018, January 17). *OpenStreetMap Wiki,* . Retrieved 09:24, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Proposed_features/Tidal_Rivers&oldid=1551116.

Quality assurance. (2019, April 15). *OpenStreetMap Wiki,* . Retrieved 07:54, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Quality_assurance&oldid=1842877.

Slippy Map. (2018, October 28). *OpenStreetMap Wiki,* . Retrieved 12:32, mai 15, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Slippy_Map&oldid=1686422.

Tag:natural=coastline. (2009, September 5). *OpenStreetMap Wiki,* . Retrieved 09:21, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Tag:natural%3Dcoastline&oldid=333953.

Tag:natural=coastline. (2016, January 14). *OpenStreetMap Wiki,* . Retrieved 09:33, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Tag:natural%3Dcoastline&oldid=1262111.

Tag:natural=coastline. (2018, November 5). *OpenStreetMap Wiki,* . Retrieved 09:23, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=Tag:natural%3Dcoastline&oldid=1689503.

TIGER fixup. (2018, October 10). *OpenStreetMap Wiki,* . Retrieved 13:45, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=TIGER_fixup&oldid=1679022.

TIGER fixup. (2018, October 10). *OpenStreetMap Wiki,* . Retrieved 08:14, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=TIGER_fixup&oldid=1679022.

Wiki. (2019, avril 14). *Wikipédia, l'encyclopédie libre*. Page consultée le 02:06, avril 14, 2019 à partir de http://fr.wikipedia.org/w/index.php?title=Wiki&oldid=158423705.

Wikidata:Property proposal. (2019, Avril 17). *Wikidata Wiki,* . Retrieved 10:30, mai 13, 2019 from https://www.wikidata.org/wiki/Wikidata:Property_proposal.

WikiProject Cadastre Français/Conditions d'utilisation. (2018, May 4). *OpenStreetMap Wiki,* . Retrieved 08:21, mai 13, 2019 from https://wiki.openstreetmap.org/w/index.php?title=WikiProject_Cadastre_Fran%C3%A7ais/Conditions_d%27utilisation&oldid=1605999.

Xerox Alto. (2018, janvier 11). *Wikipédia, l'encyclopédie libre*. Page consultée le 19:41, janvier 11, 2018 à partir de http://fr.wikipedia.org/w/index.php?title=Xerox_Alto&oldid=144393253.

Table des figures

Figure 1 Exemple d'utilisation de la méthode des ontologies pour la construction de bases de données (Hayat, 2019) 33

Figure 2 Mobilisation d'ontologies pour la construction d'une base de données cartographique pour la création d'une carte de l'île de La Réunion (Hayat, 2019) 36

Figure 3 Extraits de tags associés à trois ressources numériques, deux accessible via OpenStreetMap et une via Flickr (Hayat, 2019) 41

Figure 4 Exemple d'extrait d'un schéma conceptuel issu de la méthode de construction par ontologie (Hayat, 2019) 43

Figure 5 See et al., 2016 ; synthèse des expressions relatives à la création d'information par des particuliers 60

Figure 6 Éléments et métadonnées de la base de données OpenStreetMap (Hayat, 2019) 66

Figure 7 Différentes méthodes de contributions des adresses (Hayat, 2019) 68

Figure 8 Synthèse des composants de l'infrastructure actuelle permettant le maintien et l'utilisation de la base de données OSM 70

Figure 9 Synthèse datant de 2012 des composants de l'infrastructure permettant le maintien et l'utilisation de la base de données OSM 72

Figure 10 Extrait de la fenêtre d'installation des plugins dans JOSM 73

Figure 11 Les valeurs principales de la clé `amenity` en février 2019, capture d'écran du site taginfo 74

Figure 12 Requête sur les villes du Royaume-Uni 77

Figure 13 Requête sur la population des villes dans le monde et leur proximité à la ville de Cambridge 78

Figure 14 Accompagner le choix d'une licence par des textes simples et une conception graphique organisée 93

Figure 15 Évaluation de la compatibilité de la licence Apache sur Wikipédia 94

Figure 16: Le « succès » de la Fondation OpenStreetMap pour obtenir la mention de la source 96

Figure 17 le fichier intitulé "success.jpg" cité dans le tweet ci-contre 97

Figure 18 Tweet présentant les cartes Michelin comme fiables, capture d'écran sur Twitter 104

Figure 19 Tendance à un rapprochement de la densité des réseaux OSM et Mulinet par Denis Zielstra et Alexander Zipf (2010) 119

Figure 20 Résumé des 25 indicateurs proposés par Barron et al. (2014), figure intitulée "Overview of the iOSMAnalyzer's intrinsic Quality Indicators" par les auteurs 122

Figure 21 Diversité des résultats des analyses automatiques offerts dans Osmose (Hayat, 2019) 137

Figure 22 L'import du réseau TIGER en septembre 2007 à gauche et le réseau à la fin de l'import en janvier 2008 143

Figure 23 Challenge MapRoulette répertoriant les objets dont RB94 est le dernier modificateur, capture d'écran sur le site MapRoulette 158

Figure 24 En abscisse : la première documentation du tag ; en ordonnées : la centième utilisation dans la base de données OSM du tag 161

Figure 25 « OpenStreetMap-Wikidata Semantic Bridge », publié par Krauss en juillet 2018 177

Figure 26 Acteurs et relations entre acteurs examinés dans ce chapitre (Flora Hayat, 2019) 193

Figure 27 - Exemple de déclaration des tables avec Imposm3, ici une table (roads), ses champs (osm_id, geometry, valeur, name, tags) et sa géométrie (linestring), en marron les commentaires de l'auteure explicitant les paramètres écrits (Hayat, 2019) 197

Figure 28 Requêtes auprès du serveur par le client. (Hayat, 2019) 199

Figure 29 Exemple de paramètre de création de tuiles avec Utilery contenant une couche de routes pour la plage d'échelle 12 à 15 et s'appuyant sur un prétraitement de la donnée : le champ road_level (Hayat, 2019) 200

Figure 30 Exemple de paramètres de chargement pour une table de routes dans un SGBD PostgreSQL/GIS via Imposm3 (Hayat, 2019) 200

Figure 31 Ordre d'affichage des couches de la carte ViaMichelin (Hayat, 2019) 202

Figure 32 Première conséquence graphique du changement de source, TomTom et OSM ne présentent pas la même finesse de hiérarchisation des routes..... 203

Figure 33 Règles définies pour la cartographie automatique 204

Figure 34 Extrait d'une charte graphique retranscrite en feuille de style..... 205

Figure 35 Sélection des routes (en bleu, les tracés initiaux, en rouge les voies principales, en jaune les voies secondaires et en blanc les voies mineure selon une mise en correspondance automatique des tags d'OSM et de la charte graphique des cartes Michelin) 207

Figure 36 Extrait de la carte finale..... 208

Figure 37 Enregistrement d'une trace GPS (tracé bleu), capture d'écran du logiciel JOSM 209

Figure 38 Géolocalisation de photos (représentées par un figuré ponctuel symbolisant un appareil photographique) grâce aux relevés GPS (capture d'écran du logiciel JOSM)..... 209

Figure 39 Peser les avantages et les inconvénients pour une entreprise traditionnelle (Hayat, 2019)..... 212

Figure 40 Accès interdit au parvis Notre-Dame renseigné le 16 avril 2019 par l'inscription du tag access=no 213

Figure 41La linéarité d'une chaîne de production construite à partir de données propriétaires (Hayat, 2019)216

Figure 42 Les changements profonds méthodologiques qui accompagnent l'exploitation complète d'une source libre (hayat, 2019)..... 217

Figure 43 Un défi de développement sur OSRM lancé par Mapbox et diffusé par le compte Twitter d'OSRM 228

Figure 44 Contributions individuelles ou d'intérêts économiques à l'écosystème OSM (Hayat, 2019)..... 231

Figure 45 Les composants de l'infrastructure de service OpenStreetMap (Harry Wood, wiki OpenStreetMap, 2017)..... 236

Figure 46 En vert, les tronçons de route photographiés, plate-forme Mapillary utilisant un fond cartographique Mapbox (en vert les tronçons pour lesquels il existe des photographies, en bleu le tronçon sur lequel l'utilisateur s'est positionné, capture d'écran du site https://www.mapillary.com/, en avril 2019)..... 239

Figure 47 Exploration de la ville d'Athènes via Mapillary (capture d'écran du site https://www.mapillary.com/, en avril 2019)..... 240

Figure 48 À gauche, la représentation par Maps.me et à droite la représentation par Google Maps..... 246

Figure 49 Les meilleurs résultats du challenge #CompleteTheMap par Mapillary durant le mois d'août 2018 (capture d'écran du site https://mapillary.github.io/mapillary_greenhouse/global-challenge/q3-2018/, en avril 2019)..... 247

Figure 50 Classement en direct des contributeurs en fonction des kilomètres parcourus et du nombre de photographies chargées sur OpenStreetcam..... 248

Figure 51 Les trois panneaux de présentation au premier allumage de OpenStreetCam : gagner des points (à gauche), faire partie d'une communauté (au milieu) et la contribution est facilitée par le moyen de l'application (à droite) 248

Figure 52 Mapillary manifestant sur Twitter son avance sur OpenStreetView (OpenStreetCam aujourd'hui) 249

Figure 53 Compatibilité des licences Flickr avec la licence ODbL (Hayat, 2019)..... 261

Figure 54 Sélection de dix jeux de données parmi les quatre sources données aux licences compatibles (Hayat, 2019)..... 262

Figure 55 Exploitation de la classification du Street Art et de ses sous classes afin de construire un champ lexical associé (Hayat, 2019) 263

Figure 56 Discontinuités de classification du réseau (capture d'écran de la slippy map d'openstreetmap.org) 266

Figure 57 Résumé statistique des neuf variables et affectation des points (Hayat, 2019) 276

Figure 58 Résumés statistiques des variables emprise des espaces de promenade, emprise des bâtiments patrimoniaux et emprise des terrasses (Hayat, 2019) 277

Figure 59 Cellules qui comprennent au moins un service de Wifi, une photographie de Street Art ou une surface en eau (Hayat, 2019) 278

Figure 60 Résumés statistiques des variables hauteur moyenne du bâti, densité de voies piétonnes et densité des arrêts de bus (Hayat, 2019) 279

Figure 61 Les critères retenus en une possible répartition dans une cellule de la grille (Hayat, 2019) 280

Figure 62 La cellule théorique obtient 16 points et se positionne donc parmi les cellules les plus recommandables (Hayat, 2019)..... 281

Figure 63 Construction de la légende de la carte de synthèse des neuf critères (Hayat, 2019) 282

Figure 64 Cartographie du classement des cellules selon les neuf critères d'attractivité (Hayat, 2019) 283

Figure 65 Exemple de plan rapproché de cellules qui obtiennent un rang élevé (Hayat, 2019) 284

Figure 66 Frise chronologique d'éléments représentatifs facteurs de l'émergence du projet OSM (Hayat, 2019) 293

Figure 67 Services commerciaux de réalités augmentées à partir des données OSM et d'images satellites proposés par Mapbox 314

Figure 68 Présentation par Waze de son utilisation des données fournies par ses utilisateurs 314

Figure 69 Comparaison des recommandations d'Ostrom et l'organisation d'OSM (Hayat, 2019)..... 329

Figure 70 Candidats élus et leurs activités professionnelles et associatives, en gris quatre élus pour lesquels nous n'avons pu réunir d'informations..... 350

Table des matières

Avant-propos 1

INTRODUCTION GÉNÉRALE 3

PARTIE I : Informations géographiques à l'ère du numérique 13

Chapitre 1 : Changements de paradigme en cartographie 17

1.1. Formalisation informatique de l'information géographique 23

1.1.1. De l'observation de la réalité à la base de données géographique : décrire et prescrire une vue de la réalité 26

1.1.2. Pour l'interopérabilité des bases de données géographiques 31

1.2. L'information géographique volontaire, contributive et libre 45

1.2.1. Le Libre et les communs pour l'information géographique 46

1.2.2. Volontaire et contributif 54

1.2.3. OpenStreetMap : une base de données géographique contributive 64

1.2.4. Wikidata et Openstreetmap, des sources d'informations volontaires, contributives et libres 75

1.3. Les droits d'utilisation des données spatiales 82

1.3.1. Construction de licences de distribution libres 84

1.3.2. Acteurs et outils de communication 91

Conclusion du chapitre 1 99

Chapitre 2 : Fiabilité de la base de données OpenStreetMap 103

2.1. Qualité et crédibilité des informations géographiques volontaires et libres 108

2.1.1. Comparer OSM à une référence 115

2.1.2. Qualité d'une base de données en fonction des propriétés intrinsèques des données 120

2.1.3. La crédibilité des contributeurs 126

2.2. Du projet OpenStreetMap à l'institution OpenStreetMap ? 130

2.2.1. Fiabilisation par une pratique scientifique de la collecte et de l'édition de la base OSM 132

2.2.2. Fiabilisation par l'acquisition d'une légitimité 140

2.2.3. Vulnérabilité 152

2.3. OSM, l'incompris 159

2.3.1. Ce que le projet n'est pas 159

2.3.2. Ce que le projet peut devenir 163

Conclusion du chapitre 2 171

PARTIE I. Conclusion : OSM une source unificatrice ? 175

PARTIE II : Cartographie industrielle et communs numériques 179

Chapitre 3 : Entre publications cartographiques commerciales et productions libres d'informations géographiques 183

3.1. Les systèmes de production cartographique 194

3.1.1. Les chaînes de production 195

3.1.1.1. Une base de données cartographique issue d'une base de données folksonomiques . 195

3.1.1.2. Les tuiles vectorielles 198

3.1.1.3. Les tuiles vectorielles et les modèles de données 199

3.1.1.4. De la charte graphique au rendu en ligne 201
3.1.1.5. Productions papier 206
3.1.2 Entre nouveaux systèmes de production et assimilation des systèmes traditionnels 210
3.2. Services et produits cartographiques dérivés de l'information géographique libre 220
3.2.1. Construction d'un écosystème ouvert et spatial 222
3.2.1.1 Un ensemble de services 233
3.2.1.2. La chaîne de production de la Slippy Map d'OpenStreetMap 235
3.2.2. Les producteurs de données, utilisateurs de données, vendeurs de services 237
3.2.2.1. Des startups devenues de grandes entreprises de services 238
3.2.2.2. Des alternatives à Google Maps 242
3.2.2.3. Les relations entre les acteurs commerciaux et les communautés 247
3.2.2.4. Inciter à la contribution 251
3.3. Exploitation des services et expérimentation sur l'interopérabilité de sources ouvertes pour la cartographie touristique 253
3.3.1. Construction d'une base de données spatiales 258
3.3.1.1 Extraction, sélection et analyse des données de Flickr 262
3.3.3.2. Extraction, sélection et analyse des données d'OSM 263
3.3.3.3. Extraction et sélection des données de la Mairie de Paris et de l'APUR 267
3.3.3.4. Une base de données géographique ? 268
3.3.2. Des critères croisés 271
Conclusion du chapitre 3 285
Chapitre 4 : Quel avenir pour l'Information géographique libre ? 289
4.1. Internet, un contexte spécifique pour de nouveaux cadre de production des données 291
4.1.1. Des modes de production simultanés 294
4.1.1.1. Le minitel 296
4.1.1.2. Les protocoles techniques et sociaux de communications 299
4.1.1.3. D'un internet confidentiel au Web grand public 304
4.1.2. OSM, une construction socio-technique au sein d'Internet 306
4.1.2.1. Inventer les données spatiales et leurs outils 306
4.1.2.2. Les données spatiales en ligne 311
4.2. Production par les pairs 315
4.2.1. Communs informationnels 317
4.2.2. OSM : une organisation autour d'un bien commun façon Ostrom ? 322
Conclusion du chapitre 4 332
PARTIE II. Conclusion 333
CONCLUSION GÉNÉRALE 337
OSM comme objet à définir 337
OSM comme système mouvant : discontinuités entre les acteurs extérieurs et les objectifs initiaux du projet OSM 338
OSM comme cas d'étude du courant Libre 340
OSM comme source de données pour la cartographie 342

Annexes 347

Annexe 1 : Les contributeurs aux photos Flickr portant le *tag* Street Art, contribuées 23 mars 2018 au 23 mars 2019 à Paris 347

Annexe 2 : Les candidats et élus au bureau de la Fondation OSM de 2007 à 2018 348

BIBLIOGRAPHIE 351

Livres, articles et rapports 351

Sites Web 360

Wikis 361

Table des figures 364

Table des matières 367

Résumé 370

Résumé

Mots-clés: OpenStreetMap, Cartographie, Information géographique libre, Internet contributif, communs de la connaissance, licences, interopérabilité, système de production

La thèse porte sur l'usage cartographique à visée commerciale de la base de données libre OpenStreetMap (OSM). L'usage d'une base de données produite par des contributeurs anonymes, a priori non experts de l'information géographique (IG) et n'étant tenus à aucun engagement de qualité, suscite des interrogations. Mais dans le même temps, cette situation nouvelle révèle des enjeux d'intérêt public majeurs comme celui de l'accès à l'IG. Chacun des contributeurs s'attache à suivre des méthodes, identiques car collectivement mises au point, pour décrire des objets pérennes de l'espace public tout en collectant des informations qui les concernent dans leur pratique. Nous rappelons que la base de données présente de fortes hétérogénéités de qualité et de densité de données en fonction des types de territoire. À partir de ce constat nous expliquons que la production de cartes, prenant comme source OSM, est possible lorsque l'ensemble du système technique et social de la communauté OSM est pris en compte dans le processus de création. Nous étudions donc, l'écosystème qui soutient cette production et les modalités de diffusion de la base de données. Nos analyses nous permettent d'entrevoir soit un futur respectueux du projet, soit au contraire son aliénation. Dans le cadre d'une recherche menée durant un contrat CIFRE, nous interrogeons le changement de paradigme lié à la conception cartographique et posons la question de la réalisation menée à partir d'une base de données contributive, libre (dans sa modélisation et dans sa production) et diffusée sous une licence ouverte, tout en respectant la convention tacite entre le lecteur de la carte et son producteur suivant laquelle la sélection éditoriale n'est pas dictée par une déficience d'informations.

Summary

OpenStreetMap, Mapping, free geographic information, contributory Internet, Common knowledge, Licenses, Interoperability, Production System

This thesis focuses on the commercial cartographic use of the OpenStreetMap (OSM) free database. The use of a database produced by anonymous contributors, possibly not experts in geographical information (GI) and not bound by any quality commitment, raises questions. This new situation reveals major public interest issues such as the access to GIs. Each contributor strives to follow the same methods, developed collectively, to describe permanent objects in the public space while collecting information that is of interest in their practice. We point out that the database presents strong heterogeneity in data quality and density according to the types of territory. From this observation we explain that the production of maps, taking as source OSM, is possible when the entire technical and social system of the OSM community is taken into account in the creation process. Therefore we are therefore studying the ecosystem that supports this production and the dissemination's methods of the database. Our analyses allow us to foresee either a future that respects the project or, on the contrary, its alienation. Our research was conducted as part of a corporate contract, so we were led to question the paradigm shift related to cartographic design and the realization of a contributory database, free (in its modeling and production) and distributed under an open license, while respecting the tacit agreement between the map reader and its producer that editorial selection is not dictated by an information deficit.