



HAL
open science

OF4OSM : un méta-modèle pour structurer la folksonomie d'OpenStreetMap en une nouvelle ontologie

Anthony Hombiat

► To cite this version:

Anthony Hombiat. OF4OSM : un méta-modèle pour structurer la folksonomie d'OpenStreetMap en une nouvelle ontologie. Informatique. Université Grenoble Alpes, 2017. Français. NNT : 2017GREAM044 . tel-01683028

HAL Id: tel-01683028

<https://theses.hal.science/tel-01683028>

Submitted on 12 Jan 2018

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.

THÈSE

Pour obtenir le grade de

DOCTEUR DE LA COMMUNAUTÉ UNIVERSITÉ GRENOBLE ALPES

Spécialité : **Informatique**

Arrêté ministériel : 25 mai 2016

Présentée par

Anthony Hombiat

Thèse dirigée par **M. Jérôme Gensel**

et co-encadrée par **Mme Marlène Villanova-Oliver**

préparée au sein du **Laboratoire d'Informatique de Grenoble**

et de l'**École Doctorale Mathématiques, Sciences et Technologies de
l'Information, Informatique**

OF4OSM

Un méta-modèle pour structurer
la folksonomie d'OpenStreetMap
en une nouvelle ontologie

Thèse soutenue publiquement le **24 février 2017**,
devant le jury composé de :

M. Jérôme Euzenat

Directeur de Recherche HDR, INRIA délégation Alpes, Président

M. Christophe Cruz

Maître de Conférences HDR, Université Bourgogne Franche-Comté, Rapporteur

Mme Anne Ruas

Ingénieur des Ponts, des Eaux et des Forêts HDR, IFFSTTAR Marne-la-Vallée,
Rapporteur

Mme Nathalie Aussenac-Gilles

Directrice de Recherche HDR, CNRS délégation Midi-Pyrénées, Examinatrice

M. Guillaume Allègre

Membre de l'association OpenStreetMap France, Invité

M. Jérôme Gensel

Professeur, Université Grenoble Alpes, Directeur de thèse

Mme Marlène Villanova-Oliver

Maître de Conférences, Université Grenoble Alpes, Co-Encadrante de thèse



Remerciements

Je souhaite exprimer ma plus grande reconnaissance aux membres de mon jury de thèse qui ont accepté d'évaluer mes travaux. Je remercie d'abord le président du jury, M. Jérôme Euzenat, directeur de recherche HDR à l'Institut National de Recherche en Informatique et en Automatique (INRIA) de Grenoble.

Je remercie ensuite mes rapporteurs, M. Christophe Cruz, Maître de Conférences HDR à l'Université de Bourgogne Franche-Comté (UBFC) au Laboratoire Electronique, Informatique et Image (Le2i) et Mme Anne Ruas, Ingénieur des Ponts, des Eaux et des Forêts (IPEF) à l'Institut Français des Sciences et Technologies des Transports, de l'Aménagement et des Réseaux (IFSTTAR), dont les retours ont permis d'améliorer la qualité de ce manuscrit.

Pour avoir pris le temps d'examiner ma thèse, merci également à Mme Nathalie Aussenac-Gilles, Directrice de Recherche au Centre National de la Recherche Scientifique (CNRS) à l'Institut de Recherches en Informatique de Toulouse (IRIT). Pour l'éclairage qu'il m'a donné sur l'écosystème OpenStreetMap et sa communauté dont il est un contributeur précieux, merci à M. Guillaume Allègre, membre de l'association OpenStreetMap France.

Je tiens à exprimer mon immense gratitude à mon directeur de thèse, M. Jérôme Gensel, Professeur à l'Université Grenoble Alpes (UGA), ainsi qu'à ma co-encadrante Mme Marlène Villanova-Oliver, Maître de Conférences à l'UGA. Jérôme, merci pour la stimulation intellectuelle de tes questionnements et de tes mises en perspective. Marlène, merci pour ton recul, ta patience et surtout ton soutien.

Par ailleurs, je veux remercier les membres de l'équipe de recherche STEAMER dans laquelle j'ai eu le plaisir de travailler au cours des cinq dernières années. En premier lieu, merci à Mme Paule-Annick Davoine à la direction de l'équipe. J'insiste également pour renouveler mes sentiments d'amitié envers M. Philippe Genoud et Mme Danielle Ziebelin pour leur bienveillance depuis notre collaboration outre-atlantique à Vancouver.

Ma sympathie va bien entendu à mes collègues et amis, M. Mahfoud Boudis, M. Sylvain Bouveret, à mes camarades de bureau Mme Marie-Jeanne Natete, M. David Noël, M. Tales Paiva Nogueira, M. André Sales-Fontales, aux doctorant.e.s qui ont été de très bons conseils, Mme Betül Aydin, Mme Camille Bernard, Mme Camille Cavalière, M. Jacques Gautier, Mme Cécile Saint-Marc et Mme Mouna Snoussi, aux ingénieurs, dont la veille technologique fût d'un grand secours, M. Alexandre Audouard, M. Mohamed Amine Ghorbali, M. Thomas Leysens, M. Laurent Poulenard et M. Benoît Le Rubrus, et, enfin, aux stagiaires qui ont apportés un vent de renouveau à l'équipe Mme Camille Briend, M. Pierre-Louis Centonze, M. Oton Čopi, M. Benjamin Delhomme et Mme Clémence El Maaoui-Merzeau. Je salue également le travail d'intendance et d'assistance de Mme Latifa Boudiba, Mme Pascale Poulet, Mme Samantha Sanchez et M. Gilles Thieblemont.

Plus encore, j'aimerais souligner la contribution de M. Adel Mehiris qui a implémenté la base du prototype développé pour soutenir les modèles et méthodes élaborés dans le cadre de cette thèse. J'ai pris beaucoup de plaisir à voir se concrétiser mes travaux de recherche en la compagnie d'Adel.

Je dois aussi remercier M. Michel Ida, directeur des Open Labs ainsi que MM. Philippe Caillol et Timothée Silvestre à la direction du Minatec IDEAs Laboratory® (MIL), M. Fabrice Forest, M. Roger Pitiot, M. Éric Ismail, M. Adrien Martinière, M. Carlos Santana, M. Remi Durif et M. Aurélien Marty, Mme Marie Poulle, Mme Marie-Morgane Teirlynck et M. Nicolas Rosselot et tous les membres du MIL.

Pareillement, je rends hommage à la communauté du projet OpenStreetMap et à l'association OpenStreetMap France.

Merci à Bernard, Olivier et Sylvain pour la varappe méridienne, merci à David, Jacques, Pierre et Ugo pour les notes houblonnées sérotinales.

Un grand merci à mes colocataires et amis, Veronique, Paul-Emmanuel, Benoît et Faustine, qui m'ont supporté (au propre comme au figuré) durant ces quatre années de thèse.

Merci, enfin, à ceux à qui je dois tout le reste : Bru, Gui, Lo, Tang, Arnaud, Nico, Framb, Céd, Benj, Marty, Lyuba, Mathildou et mes parents.

Résumé

Depuis les années 2000, les technologies du Web permettent aux utilisateurs de prendre part à la production de données : les internautes du Web 2.0 sont les nouveaux capteurs de l'information. Du côté de l'Information Géographique affluent de nombreux jeux de données en provenance de plates-formes de cartographie participative telles qu'OpenStreetMap (OSM) qui a largement impulsé le phénomène de l'Information Géographique Volontaire (*Volunteered Geographic Information*, VGI). La communauté OSM représente aujourd'hui plus de trois millions de contributeurs qui alimentent une base de données géospatiales ouverte dont l'objet est de capturer une représentation du territoire mondial.

Les éléments cartographiques qui découlent de ce déluge de VGI sont caractérisés par des *tags*. Les catégorisation du contenu par tag est une méthode simple, rapide et largement répandue sur les plates-formes de *crowdsourcing*. Cependant, cette approche est un obstacle majeur pour le partage et la réutilisation de ces grands volumes d'information. En effet, ces ensembles de tags, ou *folksonomies*, sont des modèles de données beaucoup moins expressifs que les ontologies.

Dans cette thèse, nous proposons un méta-modèle pour structurer la folksonomie OSM tout en préservant sa flexibilité. Ce méta-modèle permet d'identifier différents types de tags OSM puis de les enrichir sémantiquement en une base connaissances OSM sous la forme d'une ontologie : OF4OSM. Comparativement aux travaux existants, cette ontologie présente une couverture des tags supérieure, une expressivité formelle accrue, une meilleure interconnexion avec d'autres bases de connaissances et respecte la philosophie participative du projet OpenStreetMap.

Ces travaux ouvrent la voie à la constitution d'une base de connaissances structurées des données géographiques issues de la VGI intégrée dans le Web des Données Ouvertes et Liées (*Linked OpenData*, LOD). Cette nouvelle base sera un support pour le contrôle des données au moment de l'acquisition, mais également une aide pour la restitution de l'information, ceci dans le but d'augmenter la qualité de l'Information Géographique Volontaire.

Abstract

Post-2000s web technologies have enabled users to engage in the information production process: Web 2.0 surfers are the new data sensors. Regarding Geographic Information (GI), large crowdsourced datasets emerge from the Volunteered Geographic Information (VGI) phenomenon through platforms such as OpenStreetMap (OSM). The latter involves more than three millions contributors who aim at mapping the world into an open geospatial database.

This deluge of VGI consists of spatial features that are described with tags. Content categorization with tags is typical of crowdsourcing platforms. However, this approach is also a major impediment to interoperability with other systems that could benefit from this huge amount of bottom-up data. Indeed, folksonomies are much less expressive data models than ontologies.

In this thesis, we address the issue of loose OSM metadata by proposing a meta-model for semantically lifting the OSM folksonomy while preserving the flexibility of the tagging activity. This meta-model supports the identification of different types of OSM tags and their semantification into an OSM knowledge base in the form of an ontology: OF4OSM. Comparatively to existing approaches, this ontology features a higher coverage of the OSM tags, an enhanced formal expressivity, a wider interconnection to other knowledge bases and a full compliance to the participatory philosophy of the OpenStreetMap project.

This work paves the way to the building of a knowledge base of VGI structured data that is integrated in the Linked OpenData cloud. This new database will support data quality control during the acquisition of information and assist data retrieval, in order to improve the quality of Volunteered Geographic Information.

Table des matières

Table des matières	xi
Acronymes	xiv
Liste des figures	xvii
Liste des tableaux	xix
Liste des extraits de code	xxi
1 Introduction	1
1.1 Contexte	1
1.1.1 Approche ascendante de la production d'information	1
1.1.2 Information Géographique Volontaire	3
1.2 Problématique	4
1.3 Organisation du manuscrit	6
1.3.1 Partie I : État de l'art	6
1.3.2 Partie II : Contributions	7
I État de l'art	9
2 OpenStreetMap : Principes et Limites	11
2.1 Introduction	11
2.2 Historique du projet	12
2.3 Caractéristiques principales	12
2.4 Le modèle de données OpenStreetMap	16
2.5 Système de gestion de versions	21
2.6 Les tags OpenStreetMap	23
2.7 Limites des tags OSM	25
2.8 Conclusion	27

3	Méthodes de structuration de la folksonomie OpenStreetMap	29
3.1	Introduction	29
3.2	Wiki OpenStreetMap	30
3.3	<i>Linked GeoData</i> (LGD)	34
3.4	OSMonto	37
3.5	Game With A Purpose (GWAP)	40
3.6	<i>OSM Semantic Network</i> (OSN)	41
3.7	Conclusion	42
4	Limites des méthodes de structuration de la folksonomie OpenStreetMap	45
4.1	Introduction	45
4.2	Une couverture lacunaire des tags	46
4.3	Une interconnexion incomplète	47
4.4	Une structure à l'expressivité limitée	49
4.5	Une implication faible des contributeurs OSM	52
4.6	Vers un nouveau modèle pour la sémantisation des tags OSM	52
4.7	Conclusion	54
II	Contributions	57
5	Le méta-modèle de OF4OSM	59
5.1	Introduction	59
5.2	Représentation des tags OSM dans OF4OSM	60
5.2.1	Une représentation conforme aux standards	60
5.2.2	Représentation générale d'un tag OSM	63
5.2.3	Représentation des différents types de clé	63
5.2.4	Représentation des différents types de valeur	66
5.2.5	Typologie des tags OSM	67
5.3	Une ontologie OSM plus expressive	69
5.3.1	Hiérarchie initiale clé-tag	72
5.3.2	Représentation de concepts de haut niveau	74
5.3.3	Représentation de concepts aristotéliens	78
5.4	Conclusion	81
6	Génération de l'ontologie OF4OSM	85
6.1	Introduction	85
6.2	Alimentation initiale de l'ontologie OF4OSM à partir du wiki OSM	86
6.2.1	Récupération du contenu de la page <i>Map features</i>	86

6.2.2	Alimentation initiale de OF4OSM à partir du contenu de la page <i>Map features</i>	90
6.3	Alimentation complémentaire de l'ontologie OF4OSM à partir de la base de données OSM	93
6.3.1	Récupération des tags OSM	93
6.3.2	Caractérisation des tags	97
6.3.3	Structuration des tags	102
6.4	Conclusion	107
7	Exploitation de l'ontologie OF4OSM	111
7.1	Introduction	111
7.2	Apports pour l'activité de tagging	112
7.2.1	Un étiquetage plus précis	112
7.2.2	Un étiquetage plus complet	114
7.2.3	Support pour l'identification d'erreurs de tagging	119
7.2.4	Aide à la contribution actualisée	120
7.2.5	Aide à la contribution contextualisée	122
7.3	Apports pour l'extraction de données	124
7.3.1	Une typologie des objets géographiques plus fine	124
7.3.2	Une ontologie révélatrice des pratiques émergentes	126
7.3.3	Une interopérabilité accrue	128
7.3.4	Une exploration contextualisée	128
7.4	Conclusion	130
8	Bilan et perspectives	133
8.1	Bilan des réalisations	133
8.2	Perspectives	136
	Bibliographie	139
A	Ontologie MUTO	149
B	Langage OWL	151
C	Valeurs multiples associées à la clé amenity	153
D	Ontologie AO	155
E	Extrait du code HTML de la page Web <i>Map Features</i> sur le wiki OSM	159

Acronymes

- ABox** boîte d’assertion (*Assertion Box*)
- ACE** Editeur de Classes Aristotéliennes (*Aristotelian Class Editor*)
- AKSW** groupe de recherche pour le Web Sémantique et l’Ingénierie de la Connaissance Agile (*Agile Knowledge Engineering and Semantic Web research group*)
- AO** Ontologie Aristotélienne (*Aristotelian Ontology*)
- API** interface de programmation (*Application Programming Interface*)
- BBOX** boîte englobante (*Bounding Box*)
- BRD** Contribution/Révocation/Discussion (*Bold/Revert/Discuss*)
- CC** *Creative Commons*
- CLC** *Corine Land Cover*
- CMS** système de gestion de contenu (*Content Management System*)
- CSS** *Cascading StyleSheet*
- DAG** Graphe Orienté Acyclique (*Acyclic Oriented Graph*)
- DCMI** initiative Dublin Core pour les méta-données (*Dublin Core Metadata Initiative*)
- DGI** Direction Générale des Impôts
- DnS** *Descriptions and Situations ontology*
- DOLCE** *Descriptive Ontology for Linguistic and Cognitive Engineering*
- DOM** *Document Object Model*
- DUL** Dolce+DnS Ultra Lite
- GI** Information Géographique (*Geographic Information*)
- GPS** Géo-Positionnement par Satellite (*Global Positioning System*)
- GPX** format d’échange GPS (*GPS EXchange format*)
- GWAP** jeu avec un but (*Game With A Purpose*)
- HOT** équipe humanitaire d’OSM (*Humanitarian OSM Team*)
- HTML** *HyperText Markup Language*

HTTP *HyperText Transfer Protocol*

INSEE Institut National de la Statistique et des Études Economiques

IRI *Internationalized Resource Identifier*

JAXB *Java Architecture for langage à balises extensible (Extensible Markup Language, XML) Binding*

JSON *JavaScript Object Notation*

LGD *Linked GeoData*

LOD *Données Ouvertes et Liées (Linked OpenData)*

MOAT *Meaning Of A Tag*

MUTO *ontologie unifiée de l'activité de tagging (Modular Unified Tagging Ontology)*

ODbL *Open Database License*

OF4OSM *Ontologie extraite de la Folksonomie pour OSM (Ontology from Folksonomy for OSM)*

OKF *Open Knowledge Foundation*

OSM *OpenStreetMap*

OSN *OSM Semantic Network*

OWL *Ontology Web Language*

RAM *Random Access Memory*

RDF *Resource Description Framework*

REST *Representational State Transfer*

RFC *demande de commentaires (Request For Comments)*

SIOC *Semantically-Interlinked Online Communities*

SKOS *Simple Knowledge Organization System*

SPARQL *protocole SPARQL et langage de requête RDF (SPARQL Protocol and RDF Query Language)*

TBox *boîte de terminologie (Terminology Box)*

UGC *Contenu Généré par les Utilisateurs (User Generated Content)*

URL *localisateur uniforme de ressource (Uniform Resource Locator)*

VGI *Information Géographique Volontaire (Volunteered Geographic Information)*

W3C *World Wide Web Consortium*

Xapi *OSM eXtended API*

XML *langage à balises extensible (Extensible Markup Language)*

XSD *XML Schema Definition*

Liste des figures

2.1	Cartographie du centre ville de Grenoble produite à partir des données OSM (extraction du 03/10/2016).	13
2.2	Interface graphique de l'éditeur Java JOSM.	17
2.3	Implémentation du modèle OSM avec ses 3 objets primitifs.	19
2.4	Interface graphique du site Web d'OSM donnant à voir les tags d'une entité géographique (en l'occurrence le stade des Alpes de Grenoble).	24
3.1	Extrait du guide de bonnes pratiques pour l'utilisation des tags sur le Wiki OSM.	30
3.2	Aperçu de la hiérarchie de tags issue de l'ontologie OSMonto.	38
4.1	Aperçu de la hiérarchie de tags issue de l'ontologie LGD.	50
5.1	Processus de génération de l'ontologie OF4OSM appuyé sur un méta-modèle en deux parties : un modèle de représentation des tags OSM et un modèle conceptuel.	60
5.2	Positionnement du concept <code>of4osm:OSMTag</code> au sein des vocabulaires standards pour la représentation de l'activité de <i>tagging</i>	61
5.3	Modèle de représentation d'un tag OSM.	64
5.4	Typologie des tags OSM.	70
5.5	Modèle de représentation des concepts extraits des tags OSM.	71
5.6	Hiérarchie de concepts obtenue à partir du tag <code>amenity=school</code>	73
5.7	Exemple de taxonomie générée à partir de la méthode de hiérarchisation clé-tag sur les tags associés au Château de la Commanderie.	74
5.8	Exemple de taxonomie liée au vélo générée sur la base de la hiérarchie clé-tag.	75
5.9	Exemple de taxonomie liée au vélo construite grâce à l'instance <code>CyclingActivities</code> du concept <code>of4osm:HighLevelConcept</code> pour lier des tags ayant des caractéristiques communes mais différentes clés.	76
5.10	Exemple de taxonomie liée au vélo construite grâce à la subsomption clé-tag et au concept haut-niveau <code>CyclingActivities</code>	77
5.11	Modèle simplifié de représentation d'une classe aristotélicienne.	79

5.12	Conformité du méta-modèle OF4OSM avec l'ontologie AO pour la représentation aristotélicienne de la conjonction de tags OSM.	80
5.13	Exemple de taxonomie générée à partir de la méthode aristotélicienne appliquée à la conjonction de tags OSM.	81
5.14	Exemple d'ontologie générée à partir du méta-modèle de OF4OSM.	83
6.1	Structure de la page <i>Map Features</i> du wiki OSM.	87
6.2	Alimentation initiale de l'ontologie OF4OSM à partir de page <i>Map Features</i> du wiki OSM.	89
6.3	Visualisation de l'ontologie OF4OSM après alimentation initiale dans l'éditeur Protégé.	92
6.4	Processus d'alimentation complémentaire de l'ontologie OF4OSM à partir des tags issus de la base de données OSM.	94
6.5	Implémentation dans notre prototype de la réponse renvoyée par un serveur conforme à l'API OSM.	96
6.6	Processus de caractérisation d'un tag OSM.	100
7.1	Concept de haut niveau <i>Education</i> , concept parent des objets géographiques ayant trait à l'éducation (extrait de l'ontologie OF4OSM).	113
7.2	Taxonomies issues de la classification aristotélicienne d'un fort décrit comme une attraction touristique et d'une librairie dotée d'un café (extraits de l'ontologie OF4OSM).	114
7.3	Taxonomie issue de la classification aristotélicienne d'un terrain de tennis en terre battue (extrait de l'ontologie OF4OSM).	116
7.4	Taxonomies issues de la classification aristotélicienne d'un bar diffusant de la musique Reggae et d'un bar proposant de la bière Affligem (extraits de l'ontologie OF4OSM).	118
7.5	Taxonomie issue de la classification aristotélicienne d'un <i>fablab</i> (extrait de l'ontologie OF4OSM).	119
7.6	Taxonomie issue de la classification aristotélicienne d'un bâtiment public qui abrite les bureaux d'un organisme gouvernemental (extrait de l'ontologie OF4OSM).	121
7.7	Taxonomie issue de la classification aristotélicienne d'un terrain de jeu de boule lyonnaise (extrait de l'ontologie OF4OSM).	123
7.8	Taxonomie issue de la classification aristotélicienne d'une librairie spécialisé dans les bandes dessinées et équipée d'un espace café (extrait de l'ontologie OF4OSM).	125
7.9	Taxonomies issues de la classification aristotélicienne d'une librairie équipée d'un espace café et d'un coiffeur spécialisé dans la culture punk (extraits de l'ontologie OF4OSM).	127

7.10	Taxonomies issues de la classification aristotélicienne d'une voie d'escalade et d'un bar proposant de la bière de marque La Grihète (extraits de l'ontologie OF4OSM).	129
A.1	ontologie unifiée de l'activité de <i>tagging</i> (<i>Modular Unified Tagging Ontology</i> , MUTO)	150
B.1	<i>Ontology Web Language</i> (OWL)	152
D.1	Ontologie Aristotélicienne (<i>Aristotelian Ontology</i> , AO)	156
D.2	Diagramme de classe de l'API pour Editeur de Classes Aristotéliciennes (<i>Aristotelian Class Editor</i> , ACE)	157

Liste des tableaux

2.1	Comparaison des caractéristiques des données Institut National de la Statistique et des Études Economiques (INSEE) et OSM.	15
4.1	Comparaison des modèles de données pour les tags OSM.	53
5.1	Typologie des tags OSM.	68
6.1	Bilan quantitatif des tags OSM identifiés sur la ville de Grenoble . .	101
6.2	Bilan quantitatif des concepts de l'ontologie OF4OSM après alimentations initiale et complémentaire	109
C.1	Valeurs multiples associées à la clé <code>amenity</code>	154

Liste des extraits de code

2.1	Extrait de réponse en provenance d'un serveur implémentant l'API OSM au format XML.	20
2.2	1 ^{er} changeset de la base de données OSM.	21
2.3	Modifications apportées à la base de données OSM lors du 1 ^{er} changeset.	22
3.1	Représentation de l'église Notre-Dame de Dresde au format XML par un serveur implémentant l'API OSM.	35
3.2	Triplets représentant l'église Notre-Dame de Dresde dans l'ontologie LGD.	36
6.1	Algorithme d'alimentation initiale de l'ontologie OF4OSM à partir de la page <i>Map Features</i> du wiki OSM.	90
6.2	Extrait de la réponse du serveur implémentant l'API OSM représentant des objets géographiques situés sur la ville de Grenoble	95
6.3	Algorithme de caractérisation des clés des tags OSM.	98
6.4	Algorithme de caractérisation des valeurs des tags OSM.	99
6.5	Algorithme de création de la hiérarchie de concepts issue des tags OSM et de leur clé.	103
6.6	Algorithme pour la génération d'arrangements de k éléments parmi n.	105
6.7	Algorithme pour la génération d'arrangements de 2 à n éléments parmi n.	105
6.8	Algorithme pour la génération d'une hiérarchie de concepts issus de la conjonction de tags associés à un même objet géographique.	107

Chapitre 1

Introduction

1.1 Contexte

1.1.1 Approche ascendante de la production d'information

À l'heure où l'Internet devient de plus en plus ubiquitaire¹, la posture des citoyens face à l'information connaît de profonds bouleversements. Auparavant simples consommateurs de données, les internautes sont aujourd'hui amenés à prendre part à la production et à l'interprétation de l'information. Selon M. Goodchild, les citoyens qui se trouvent du « bon » côté de la fracture numérique [Weidmann et al., 2016] sont les nouveaux capteurs de l'information (*citizen sensors* [Goodchild, 2007]). Les progrès dans les technologies de l'information et de la communication, l'accès à l'Internet à grande échelle et à haut débit, l'augmentation de la puissance des ordinateurs permettant de supporter la visualisation dynamique, ainsi que les technologies du Web 2.0 (blogs, wikis, réseaux sociaux, syndication de contenu, architecture de type REST, *etc.*) constituent un nouveau paradigme grâce auquel les utilisateurs sont impliqués dans la construction des contenus en ligne [O'Reilly, 2005].

Alors que la barrière entre la production et la consommation de l'information tombe, A. Bruns invente le terme "*produsage*" pour parler du processus collaboratif et en continu de construction et d'extension de contenus existants dans le but d'une amélioration future [Bruns, 2006]. Les sites Web orientés vers la co-construction des contenus tels que les wikis affluent et bouleversent les enjeux de pouvoir relatifs à la

1. On comptait un milliard d'utilisateurs du réseau Internet en 2005, deux milliards en 2010 et finalement plus de trois milliards aujourd'hui. <http://www.internetlivestats.com/internet-users>, consulté le 15/11/2016.

production de l'information [Yang et al., 2008; Johnson and Sieber, 2013]. Le plus célèbre d'entre eux, l'encyclopédie libre Wikipédia², compte aujourd'hui près de 30 millions de contributeurs et plus de 40 millions de pages³. Au moyen d'outils tels que Wikipédia, et à la différence du contenu personnel assisté par les blogs et les réseaux sociaux, s'opère une "mise en données du monde" [Cukier, 2013] par les internautes. Cette science citoyenne (*citizen science* [Silvertown, 2009]) est le fruit d'une approche ascendante dans le processus de production de l'information qui est à contre-courant de l'approche descendante adoptée par les institutions.

Les données citoyennes possèdent des caractéristiques radicalement différentes de celles qui sont propres aux données institutionnelles. Tout d'abord, la donnée citoyenne est ouverte à la contribution sur la base du volontariat. Ainsi, toute personne équipée d'un accès à Internet peut potentiellement ajouter, modifier ou supprimer de l'information. De plus, le processus de collecte des données s'effectue en continu. En effet, les *producers* peuvent alimenter le flux d'information à tout moment, ce qui permet de recueillir une information actualisée. L'ouverture au plus grand nombre et sans interruption, a pour effet de générer des flux massifs de données.

Pour encadrer la diffusion de ces grandes quantités de données citoyennes, des outils juridiques adaptés voient le jour : les licences dites "ouvertes". Parmi elles, les licences *Creative Commons* (CC)⁴ essaient de trouver un équilibre entre un carcan de type "tous droits réservés", qui régit habituellement la propriété intellectuelle, et celui de type "aucun droit réservé", que l'on retrouve dans le domaine public [Loenen, 2012]. Ce cadre juridique stimule l'effervescence que connaît aujourd'hui l'information citoyenne en facilitant son accès et sa réutilisation. Largement impulsé par la réorganisation des outils du Web autour des utilisateurs et de la profusion de données citoyennes qui en résulte, le mouvement pour l'ouverture des données (OpenData) encourage tout fournisseur de données (organisation publique, privée ou simple citoyen) à autoriser l'accès et la réutilisation sans restriction des données qu'il produit via ce type de licences [Loenen, 2012].

La réutilisation et l'interopérabilité de ces grandes quantités de données citoyennes ouvertes sont galvanisées par des méta-données, elles-aussi, *crowdsourcées*. En effet, la description des données ouvertes se fait au moyen d'un système d'étiquetage. Les étiquettes, ou *tags*, telles que les *hashtags* utilisés sur le site Web Twitter.com⁵, permettent aux internautes de catégoriser du contenu. Les tags sont des mots-clés en texte libre qui forment des modèles de représentation des connais-

2. <https://www.wikipedia.org/>

3. <https://en.wikipedia.org/wiki/Wikipedia:Statistics>, consulté le 15/11/2016

4. <http://creativecommons.org/licenses/>

5. <https://support.twitter.com/articles/231414>

sances appelées *folksonomies*. La facilité d'utilisation de ce type de classification, qui contraste avec les ontologies du Web sémantique [Shirky, 2005; ASIS&T, 2015], permet, une nouvelle fois, de mettre à contribution l'intelligence collective [O'Reilly, 2005].

1.1.2 Information Géographique Volontaire

L'approche ascendante de la production de l'information affecte tous les domaines de connaissances et, en particulier, celui de l'Information Géographique. En effet, au cours de la dernière décennie, la participation des citoyens aux processus de collecte et d'édition de données géographiques est en plein essor. Héritier du mouvement de l'OpenData, ce nouveau phénomène est appelé Information Géographique Volontaire (*Volunteered Geographic Information*, VGI). Le terme VGI est inventé par M. Goodchild dans un article de 2007, intitulé *Citizen as Sensors : the World of Volunteered Geography (Les citoyens comme capteurs : le Monde de la Géographie Participative)* [Goodchild, 2007].

La VGI s'inscrit dans le cadre plus général du Contenu Généré par les Utilisateurs (*User Generated Content*, UGC). Elle recouvre les différentes formes de production et de traitement de données spatialement référencées par les citoyens. L'information ainsi produite est rendue accessible et réutilisable librement par tous de façon à faciliter l'échange et la collaboration entre les différents agents. Dans cet article, M. Goodchild insiste sur la valeur des contributions de non-experts. En effet, les contributeurs, qui n'ont pas, *a priori*, de connaissances générales dans le domaine de l'Information Géographique, ont tendance à décrire les zones qui leur sont familières. Ainsi, par l'intermédiaire des logiciels de VGI, ils cartographient l'environnement dont ils sont coutumiers (leur quartier, leur lieu de travail, *etc.*) à l'échelle de leurs déplacements quotidiens. Du même coup, la représentation qu'ils donnent de ce territoire exprime une expérience terrain que les moyens limités d'une expertise (temps, ressources humaines) peinent à saisir, en particulier à une large échelle.

Les applications qui mettent la géographie participative à l'honneur se développent sur la Toile. Parmi elles se dégage un acteur majeur dans le domaine de la VGI : le projet OpenStreetMap (OSM). Ce projet, qui voit le jour en 2004, compte aujourd'hui plus de trois millions d'utilisateurs⁶ qui contribuent à alimenter la plus importante base de données géographiques ouverte et libre du monde. En l'espace d'une dizaine d'années, plus de cinq milliards de points GPS ont été collectés⁷ dans

6. <http://osmstats.neis-one.org/>, consulté le 15/11/2016

7. http://www.openstreetmap.org/stats/data_stats.html, consulté le 15/11/2016

une base de données spatialement référencées. À l’instar des données de Wikipédia, cette information est régie sous une licence ouverte : la licence *Open Database License* (ODbL) 1.0⁸. Ce cadre juridique autorise le partage à l’identique de l’information contenue dans une base de données. La libération de l’Information Géographique la rend accessible à tous et de nombreuses applications en ligne fondent leur service sur la réexploitation de la base de données OSM⁹.

Tout comme les données issues de l’OpenData au sens large, la VGI est produite par des contributeurs qui sont le plus souvent non-initiés à la géographie. De plus, les outils qui appuient la production de VGI privilégient l’accessibilité au plus grand nombre sur le contrôle de la qualité des contributions. Par ailleurs, les *producers* sont animés par différents types de motivations, certaines constructives (altruisme, stimulation intellectuelle, expression personnelle, reconnaissance sociale, *etc.*), d’autres nuisibles (malice, calcul politique, intention criminelle, *etc.*) [Coleman et al., 2009; Coleman, 2010]. Par conséquent, la VGI soulève certains problèmes de qualité parmi lesquels le manque d’expertise [McCall and Minang, 2005], l’incomplétude [Zielstra and Zipf, 2010; Neis et al., 2011; Dorn et al., 2015], l’incohérence [Sehra, 2014], l’imprécision [Canavosio-Zuzelski et al., 2013], l’inadéquation de son méta-modèle [Haklay, 2010] et de ses méta-données [Jilani et al., 2014]. C’est ce dernier aspect qui est au coeur de la problématique de recherche pour laquelle nous proposons des réponses dans ce manuscrit de thèse.

1.2 Problématique

Le projet OpenStreetMap est aujourd’hui le fer de lance de l’Information Géographique Volontaire. Tout à la fois les volumes massifs de données qu’il collecte, la disponibilité de son information, la pérennité du projet et l’ampleur de sa communauté nous ont conduits à le mettre au centre de nos recherches sur la qualité des données géographiques citoyennes.

La base de données OSM possède un modèle de données basé sur trois types primitifs : le noeud (*node*), un objet ponctuel représenté par une longitude et une latitude ; la ligne (*way*), constituée d’un ensemble de points et la relation (*relation*) qui est une liste ordonnée de noeuds, lignes ou d’autres relations¹⁰. Chaque objet cartographié correspond donc à l’un de ces trois types. Pour décrire les objets géographiques collectés dans la base de données OSM, les contributeurs utilisent des

8. <http://opendatacommons.org/licenses/odbl/>

9. http://wiki.openstreetmap.org/wiki/List_of_OSM_based_Services

10. <http://wiki.openstreetmap.org/wiki/Elements>

mots-clés aussi appelés *tags*. Les *tags* OSM sont des valeurs attributaires, sous la forme d'une paire clé/valeur, qui permettent à l'utilisateur de décrire chaque élément cartographié. La clé et la valeur de chaque tag sont définies en texte libre par les contributeurs. L'ensemble de ces tags constitue la folksonomie propre à OSM. À titre d'exemple, pour décrire une école dans la base de données OSM, un contributeur doit 1) dessiner ses limites géographiques sur la carte, 2) lui associer le tag `amenity=school` (le guide de bonnes pratiques OSM préconise l'anglais¹¹).

À la différence des ontologies, qui définissent les concepts et relations du Web sémantique réalisées par les experts du domaine [Bishr and Kuhn, 2007], les folksonomies sont, quant à elles, co-construites par les internautes. Elles sont "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" [Wal, 2005]. Les contributeurs ne les utilisent pas tant comme un moyen de faire des catégories que comme une façon de relier des éléments entre eux, et de leur attribuer une signification qui fait sens à leurs yeux [Wal, 2005]. Cette méthode de catégorisation par étiquetage (*tagging*) est très flexible puisque l'utilisateur ne rencontre aucune contrainte lorsqu'il crée ou modifie un tag.

Il est patent que l'utilisation de folksonomies comporte de sérieux inconvénients tels que les erreurs typographiques, la redondance conceptuelle et le manque d'expressivité formelle par rapport à d'autres types de classification tels que les thesauri (synonymie), les taxonomies (subsumption), ou encore les ontologies (logique du premier ordre) [Weller, 2007]. Par ailleurs, l'usage des tags peut être ambigu : S. Golder identifie sept fonctions différentes des tags, chacune liée à une intention différente de l'utilisateur [Golder, 2006] : certains tags servent à caractériser le sujet de la ressource taggée, d'autres à reconnaître son type, d'autres encore à identifier son créateur. La réutilisation de cette information devient alors d'autant plus difficile qu'elle repose sur une incertitude quant à ce qu'a voulu exprimer le tagger. Pour autant, s'en priver serait nettement plus préjudiciable puisque la description de l'information perdrait, dès lors, son caractère évolutif [Ho and Rajabifard, 2010]. Partant de ce constat, nous avons tenté, au cours de nos travaux de thèse, d'apporter des réponses à la problématique suivante : comment structurer la folksonomie OSM pour augmenter son pouvoir sémantique sans nuire à sa flexibilité ?

Plusieurs réponses aux faiblesses de la folksonomie OSM ont été proposées [Auer et al., 2009; Codescu et al., 2011; Baglatzi et al., 2012; Ballatore et al., 2013]. Dans ce manuscrit de thèse, nous présentons un état de l'art de ces travaux afin d'identifier les avantages et les limites dans le but de proposer des modèles et des

11. http://wiki.openstreetmap.org/wiki/Map_Features

méthodes pour la génération d'une ontologie de la base de connaissances OSM. Cette ontologie, que nous appelons OF4OSM, est produite à partir des tags qui servent à caractériser les objets géographiques collectés dans la base de données OSM. L'ontologie OF4OSM doit servir à améliorer la qualité des données du projet OSM, mais également à en faciliter l'extraction. La sémantique de la folksonomie OSM sera exploitée et augmentée dans l'ontologie OF4OSM, tout en préservant la flexibilité de l'activité de tagging.

1.3 Organisation du manuscrit

Ce manuscrit de thèse s'articule autour de deux parties. La partie I propose un état de l'art du domaine dans lequel s'inscrivent nos travaux organisé en trois chapitres. La seconde partie, également composée de trois chapitres, est dédiée à l'exposé de nos propositions pour pallier les limites des approches analysées en partie I. Ces travaux s'inscrivent dans le prolongement des recherches publiées dans les articles [Hombiat et al., 2015] et [Hombiat et al., 2016].

1.3.1 Partie I : État de l'art

Le chapitre 2 présente le projet OpenStreetMap. Dans un premier temps, nous dressons un historique du projet, depuis sa genèse en 2004 jusqu'au succès qu'il connaît aujourd'hui. En second lieu, nous introduisons ses caractéristiques principales. Puis, nous examinons en détail le modèle qui structure sa base de données géographiques. Par la suite, nous expliquons les principes de son système de gestion de versions. Enfin, nous abordons la technique de classification des objets cartographiés par *tags* ainsi que ses limites.

Le chapitre 3 est consacré à l'état de l'art des efforts pour la sémantisation des tags OSM. En premier lieu, le guide de bonnes pratiques pour l'utilisation des tags OSM est introduit. Ensuite, les ontologies *Linked GeoData* et *OSMonto* sont présentées. Puis, nous exposons les principes d'une méthode d'alignement semi-automatique sur une ontologie haut-niveau qui repose sur une approche "ludique" (*game with a purpose*). Enfin, ce chapitre s'achève avec la présentation du *OSM Semantic Network* qui constitue le modèle sémantique basé sur les tags OSM le plus abouti.

Les limites de ces différentes approches sont identifiées dans le chapitre 4 qui clot la première partie du présent manuscrit. La focale est mise sur quatre aspects

qui nous semblent cruciaux pour la construction d'une ontologie des connaissances OSM à partir de ces tags : 1) l'inclusion des tags dans le modèle proposé, 2) l'interopérabilité du modèle avec d'autres bases de connaissances, 3) l'expressivité formelle de la structure dans laquelle sont représentés les tags, 4) l'aspect participatif, cher au projet OSM. Finalement, les cinq modèles de structuration de la folksonomie OSM abordés dans le chapitre 3 sont analysés à l'aune de ces quatre critères afin d'identifier les limites auxquelles nous proposons des réponses dans la partie II de cette thèse.

1.3.2 Partie II : Contributions

Le premier chapitre de cette partie II, le chapitre 5, décrit en détail le méta-modèle de l'ontologie OF4OSM. Ce modèle est composé de deux parties distinctes. La première consiste en un modèle de représentation des tags OSM et de leur composantes, la clé et la valeur. La seconde est le modèle qui supporte la base de connaissances OSM en elle-même. Ce modèle est organisé en trois volets : 1) hiérarchisation initiale entre clés et tags, 2) factorisation de caractéristiques communes entre tags, 3) classification aristotélicienne des tags utilisés conjointement.

La génération de l'ontologie des tags OSM est expliquée dans le chapitre 6. Dans un premier temps, nous présentons les méthodes et outils pour l'alimentation initiale de l'ontologie OF4OSM à partir de la page *Map Features* du wiki OSM qui répertorie les tags consensuels acceptés par la communauté OSM. Dans un second temps, nous introduisons les méthodes et outils pour l'alimentation complémentaire de l'ontologie OF4OSM au travers d'un processus divisé en trois étapes : 1) extraction des données OSM via une API dédiée, 2) caractérisation des tags OSM sur la base de la typologie précédemment évoquée, 3) création de l'ontologie OF4OSM à partir des tags OSM caractérisés.

Le chapitre 7 met en avant les bénéfices tirés de l'exploitation de l'Ontologie extraite de la Folksonomie pour OSM (*Ontology from Folksonomy for OSM*, OF4OSM) qui résulte des différentes phases exposées au chapitre 6. D'une part, nous abordons les atouts de l'utilisation de OF4OSM pour l'amélioration de la qualité des données OSM au moment de l'acquisition. De l'autre, nous exposons les apports de cette nouvelle ontologie en ce qui concerne l'extraction de l'information collectée dans la base de données OSM.

Le dernier chapitre de ce mémoire de thèse, le chapitre 8, dresse le bilan des contributions exposées dans la partie II et propose des perspectives à nos travaux.

Première partie

État de l'art

Chapitre 2

OpenStreetMap : Principes et Limites

2.1 Introduction

Nous avons vu dans le chapitre précédent comment la rencontre d'évolutions technologiques dans le domaine de la géomatique et d'un désir grandissant des citoyens de prendre part au processus de production de l'Information Géographique (*Geographic Information*, GI) a engendré un phénomène en pleine effervescence : l'Information Géographique Volontaire (*Volunteered Geographic Information*, VGI). Celle-ci est aujourd'hui largement impulsée par un acteur qui poursuit la même approche ascendante que l'encyclopédie libre et collaborative Wikipédia¹ dans le domaine de la géographie : il s'agit du projet OpenStreetMap (OSM).

Dans ce chapitre, nous présentons les principes du projet OpenStreetMap que notre contribution se propose d'enrichir. Dans un premier temps, nous dressons un historique du projet, depuis sa genèse en 2004 jusqu'au succès qu'il connaît aujourd'hui (section 2.2). En second lieu, nous introduisons ses caractéristiques principales (section 2.3). Puis, nous examinons en détail le modèle qui structure sa base de données géographiques (section 2.4). Par la suite, nous expliquons les principes de son système de gestion de versions (section 2.5). Enfin, nous analysons la technique de classification des objets cartographiés par *tags* (section 2.6) ainsi que ses limites (section 2.7).

1. <https://www.wikipedia.org/>

2.2 Historique du projet

En réaction à la privatisation des données collectées par l'agence cartographique publique de son pays, l'Ordnance Survey², Steve Coast décide de lancer un équivalent à Wikipédia pour l'Information Géographique. En 2004, le projet OpenStreetMap³ voit le jour. Il s'agit d'une base de données spatiales ouverte et gratuite dont l'alimentation est rendue accessible à tous les internautes. En tant que plate-forme de cartographie participative, OpenStreetMap s'adresse avant tout à des non-experts. Comme tout projet participatif, il s'agit de mobiliser l'intelligence collective des internautes chère à T. O'Reilly [O'Reilly, 2005] et de tirer parti de la loi de Linus selon laquelle "avec suffisamment d'yeux, tous les bugs sautent aux yeux" [Raymond, 1999]. Contrairement aux approches de production de l'information ascendantes adoptées par les fournisseurs de données institutionnels, OSM donne la possibilité aux profanes de décrire l'environnement qui leur est familier à l'échelle du bâti.

Dans un premier temps, la base de données OSM ne couvre que le territoire du Royaume-Uni. Puis, deux ans plus tard, la couverture de la carte OSM s'étend à l'ensemble du globe. Début 2013, un million de contributeurs participaient à la cartographie de la Terre; on en compte plus du triple aujourd'hui⁴. Depuis son ouverture internationale courant 2006, l'acquisition de nouvelles données grâce à cette armée de volontaires a abouti à l'alimentation d'une base de données extrêmement riche : plus de 500 000 points de cheminement GPS (*track points*) ont été collectés, 3 millions d'éléments géographiques ponctuels, et 300 000 lignes et polygones⁵. Après 12 ans d'existence, le projet OpenStreetMap est devenu le fer de lance de l'Information Géographique Volontaire. Le site Web principal du projet, <http://www.openstreetmap.org/>, donne un aperçu des informations provenant de cette gigantesque base de données (la version formatée en XML représentait plus de 600GB début 2016⁶) au moyen de la carte interactive illustrée en figure 2.1.

2.3 Caractéristiques principales

L'information collectée dans la base de données OSM présente un niveau de granularité très fin. En effet, les utilisateurs contribuent à construire une représentation de l'espace à l'échelle du bâtiment ou de l'infrastructure routière. Cette représenta-

2. <https://www.ordnancesurvey.co.uk/>

3. <http://www.openstreetmap.org/>

4. <http://osmstats.neis-one.org/>, consulté le 15/11/2016

5. <http://wiki.openstreetmap.org/wiki/Stats>, consulté le 15/11/2016

6. <http://wiki.openstreetmap.org/wiki/Planet.osm>



FIGURE 2.1 – Cartographie du centre ville de Grenoble produite à partir des données OSM (extraction du 03/10/2016).

tion à échelle humaine s'appuie sur des infrastructures familières aux contributeurs. Ces derniers n'ont pas l'expertise des sciences du territoire. Cependant, ce sont ici des lieux et objets du quotidien qui sont cartographiés. Par conséquent, les utilisateurs les saisissent d'autant mieux qu'ils en sont les créateurs. Le niveau de granularité fin de l'information de la base OSM présente donc le double avantage d'être proche du citoyen et d'être imprégné d'un point de vue d'expert terrain. Par ailleurs, ces données infra-communales peuvent être collectées à n'importe quel moment là où les fournisseurs de données classiques proposent des mises à jour ponctuelles à des fréquences de l'ordre de l'année. Le caractère continu du processus d'acquisition garantit une fraîcheur des données optimale et permet de suivre l'évolution du territoire au plus près. En ce qui concerne la couverture des données collectées dans OSM, les auteurs de [Haklay, 2010] soulignent le fait que les zones urbaines sont mieux prises en considération que les zones rurales.

De plus, afin d'encourager le partage de cette information citoyenne, son créateur inscrit dès l'origine OSM dans le mouvement de l'OpenData - dont il deviendra rapidement le projet phare - qui réclame plus de transparence vis-à-vis des organismes de production de l'information (tel que l'Institut National de la Statistique et des Études Economiques (INSEE)⁷ en France). En ce sens, à l'origine, la base de données OSM était encadrée par la licence ouverte *Creative Commons* (CC) CC-by-SA (<https://creativecommons.org/licenses/by-sa/2.0/>). Cependant, cette dernière était lacunaire et fut abandonnée car elle n'avait pas été conçue spécifi-

7. <http://www.insee.fr/en/>

quement pour encadrer l'information cartographique (elle est notamment évasive quant à la protection des données telles que les géométries des objets géographiques et les tags qui y sont associés^{8 9}). La migration des données sous la nouvelle licence ne s'est pas faite sans anicroche puisque les informations appartenant à des contributeurs OSM n'ayant pas accepté la licence ODbL ont dû être supprimées de la base. Ces données ne représentaient toutefois que 1% de l'ensemble des données en base¹⁰. Depuis septembre 2012, l'utilisation de la base de données OSM est régie par la licence *Open Database License* (ODbL)¹¹ dont le cadre juridique est défini par l'un des organismes les plus influents en matière d'ouverture des données, l'*Open Knowledge Foundation* (OKF)¹². En outre, cette licence autorise tout citoyen à réutiliser les données OSM dans la mesure où il ne modifie pas leur contenu et fait référence au producteur de données. Ce cadre encourage la réutilisation de l'Information Géographique Volontaire dans d'autres projets.

Le tableau 2.1 synthétise l'ensemble des caractéristiques d'OSM en les mettant en regard des données fournies par l'INSEE, principal organisme de production de statistiques publiques en France. Le premier est l'acteur le plus important du phénomène de la VGI, tandis que le second est un fournisseur de données institutionnelles représentatif du processus classique de production de l'information. Ces deux approches sont analysées à l'aune des différents aspects évoqués plus haut : production (nombre d'agents producteurs de l'information et type d'expertise), collecte (finesse du maillage, couverture des données et fréquence de mise à jour), et disponibilité (type de licence et coût).

Nous l'avons vu, le projet OpenStreetMap s'inscrit dans le phénomène de la VGI et adopte une approche radicalement différente de celle des producteurs de données classiques. Néanmoins, il bénéficie également de l'autorité de données institutionnelles qui rejoignent elles-aussi le mouvement de l'OpenData et, en particulier, les jeux de données mis à disposition sur les plates-formes en ligne du type `data.gov.*`. Ces dernières ont pour but de promouvoir la transparence des données gouvernementales et d'encourager leur (ré)utilisation. La déclinaison française de cette initiative (américaine à la base, `www.data.gov`¹³) est le site Web `data.gouv.fr`¹⁴.

8. http://wiki.osmfoundation.org/w/index.php?title=Licence/Historic/We_Are_Changing_The_License

9. http://wiki.osmfoundation.org/w/index.php?title=Licence_and_Legal_FAQ/Why_CC_BY-SA_is_Unsuitable

10. Pour plus d'informations sur le changement de licence des données OSM, voir <http://www.openstreetmap.fr/openstreetmap-odbl>

11. <http://opendatacommons.org/licenses/odbl/1.0/>

12. <https://okfn.org/>

13. <https://www.data.gov/>

14. <http://www.data.gouv.fr/>

TABLE 2.1 – Comparaison des caractéristiques des données INSEE et OSM.

	Caractéristique	INSEE	OSM
Production	Nombre	Bas	Haut
	Expertise	Expertise du domaine	Expérience terrain
Collecte	Maillage	National à Communal	Infracommunal
	Couverture	Complète	Partielle (espaces urbains)
	Fréquence de mise à jour	Annuelle à mensuelle	Continue
Disponibilité	Licence	Principalement privée	Publique
	Coût	Elevé	Gratuit

Des imports des données mises à disposition sur ces sites Web ont permis d'enrichir significativement la base OSM, lui conférant, par là-même, un peu de leur crédit. Un catalogue des imports de données en provenance d'institutions est disponible sur le wiki du projet¹⁵. A titre d'exemple, en 2009, les données du cadastre français¹⁶ fournies par la Direction Générale des Impôts (DGI) ont été ajoutées à la base de données OSM¹⁷. La même année, la caractérisation biophysique des sols via la nomenclature *Corine Land Cover* (CLC)¹⁸ par le ministère de l'environnement a elle-aussi été importée¹⁹. Dernier ajout en date, début 2016, les établissements de tourisme accessibles en chaise roulante publiés par le Ministère de l'Economie, de l'Industrie et du Numérique²⁰. Les volontaires intéressés par l'import de données institutionnelles ont également établi une liste des sources de données potentielles disponible sur le Wiki OSM²¹.

Finalement, on peut aujourd'hui apprécier le succès d'OSM à travers les projets qui gravitent autour de lui. Le wiki OSM en dresse la liste non exhaustive²². Entre autres, on trouve des applications ayant trait au sport telles que OpenCycleMap²³ qui se propose de mettre en exergue les éléments de la base de données OSM en rapport avec l'activité du cyclisme. D'autres exploitent les informations OSM dans un but artistique comme le site Web `prettymaps.stamen.com`²⁴ qui croise les don-

15. <http://wiki.openstreetmap.org/wiki/Import/Catalogue>

16. <http://www.cadastre.gouv.fr/>

17. http://wiki.openstreetmap.org/wiki/WikiProject_France/Cadastre

18. https://fr.wikipedia.org/wiki/Corine_Land_Cover

19. http://wiki.openstreetmap.org/wiki/WikiProject_Corine_Land_Cover

20. [http://wiki.openstreetmap.org/wiki/Les_Etablissements_\"Tourisme_&_Handicap\"](http://wiki.openstreetmap.org/wiki/Les_Etablissements_\)

21. http://wiki.openstreetmap.org/wiki/Potential_Datasources

22. http://wiki.openstreetmap.org/wiki/List_of_OSM-based_services

23. <http://www.opencyclemap.org/>

24. <http://prettymaps.stamen.com/>

nées produites par plusieurs sites Web communautaires (Flickr²⁵, Natural Earth²⁶ et OSM) et applique un style graphique qui donne à voir le territoire sous une perspective esthétique. Le projet OSRM²⁷ s'appuie sur le réseau routier d'OSM pour calculer des itinéraires. Enfin, l'équipe humanitaire d'OSM (*Humanitarian OSM Team*, HOT)²⁸, tire parti de l'aspect collaboratif et de l'adaptabilité de la carte OSM pour cartographier des zones dévastées par des catastrophes naturelles, et, en particulier, les accès afin de venir rapidement en aide aux victimes.

2.4 Le modèle de données OpenStreetMap

Pour contribuer à OSM, il suffit de créer un compte (adresse mail et mot de passe) pour avoir la possibilité de modifier ou de supprimer de l'information provenant de la base de données OSM. L'ajout d'un nouvel objet en base consiste en deux étapes : 1) la matérialisation de son empreinte géographique sous la forme d'un point, d'une ligne, d'un polygone ou d'un multipolygone, puis 2) la caractérisation de l'objet à l'aide d'une ou plusieurs étiquettes (*tags*). Par exemple, pour représenter la mairie de la ville de Grenoble, il convient d'en dessiner les limites spatiales (généralement sur un fond de carte constitué de photos aériennes mis à disposition par un éditeur OSM) puis de lui associer les tags `name=MairiedeGrenoble` et `amenity=townhall`.

L'empreinte spatiale d'une entité peut également se faire par l'ajout d'une trace GPS (format format d'échange GPS (*GPS EXchange format*, GPX)) dans la base de données OSM. La base de données OSM est accessible en mode édition directement depuis le site Web du projet, mais il existe plusieurs éditeurs disponibles en application bureau. Une liste non exhaustive de ces éditeurs OSM est mise à disposition sur le wiki d'OSM²⁹. Les plus populaires sont iD³⁰, utilisé sur le site Web du projet OpenStreetMap, Potlatch2³¹, le prédécesseur de iD, et JOSM³², l'application bureau pour une édition avancée de la base de données OSM. L'interface graphique de JOSM est illustrée en figure 2.2.

Le modèle de données qui permet de représenter les objets géographiques dans OSM s'appuie sur 3 objets primitifs³³ :

25. <http://code.flickr.net/2008/10/30/the-shape-of-alpha/>

26. <http://www.naturalearthdata.com/>

27. <http://map.project-osrm.org/>

28. <https://hotosm.org/>

29. <http://wiki.openstreetmap.org/wiki/Editors>

30. <http://ideditor.com/>

31. http://wiki.openstreetmap.org/wiki/Potlatch_2

32. <https://josm.openstreetmap.de/>

33. <http://wiki.openstreetmap.org/wiki/Elements>

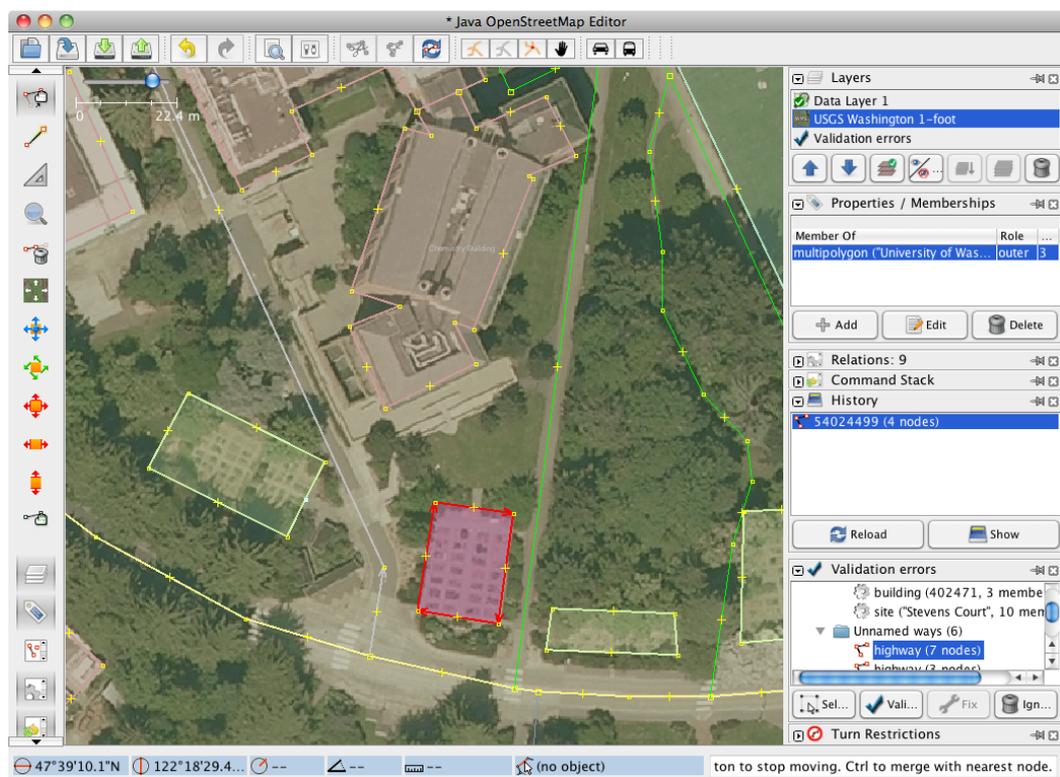


FIGURE 2.2 – Interface graphique de l'éditeur Java JOSM.

Le **noeud** (*node*) représente un point spécifique à la surface du globe, identifié par une latitude et une longitude. Les noeuds permettent de représenter des objets ponctuels comme un banc, un feu de circulation ou encore une fontaine d'eau potable. Les noeuds servent également à définir la forme d'une ligne polygonale.

La **ligne** (*way*) est une liste ordonnée de 2 à 2000 noeuds décrivant une ligne polygonale (ou polyligne). Cet élément peut servir à décrire des objets linéaires tels que des routes ou des cours d'eau. La polyligne peut également représenter les limites d'une zone telles que l'empreinte spatiale d'un bâtiment, d'une forêt ou encore d'une entité spatiale administrative (la ville de Grenoble). On parle alors de ligne fermée (*closed way*). Dans ce cas précis, le premier noeud de la liste est le même que le dernier. Lorsque l'objet géographique ne peut être représenté par 2000 noeuds ou moins, il faut recourir à une relation multipolygonale.

La **relation** (*relation*) est une structure de données permettant d'exprimer une relation entre au moins 2 des éléments précédemment décrits (le noeud et la polyligne) ou entre deux relations. La plus fréquente de ces relations est la relation polygonale qui construit un polygone à partir de plusieurs lignes fer-

mées³⁴. Il existe d'autres types de relation tels que les lignes de transport en commun (*route*), les voies navigables (*waterway*) ou encore les limites administratives (*boundary*). Le Wiki OSM repertorie l'ensemble des types de relation possibles³⁵.

Pour chacun de ces éléments, 6 attributs sont enregistrés en base de données :

id L'identifiant unique de l'objet.

uid L'identifiant unique du contributeur OSM ayant modifié pour la dernière fois l'objet.

timestamp Le moment auquel a eu lieu la dernière modification de l'objet. Cet horodatage est conforme au standard du *World Wide Web Consortium* (W3C)³⁶.

visible Indique si l'objet est toujours d'actualité (**visible=true**) ou bien s'il est désormais supprimé de la base de données (**visible=false**) et s'il ne doit être retourné que dans les résultats de requêtes à portée historique.

version La version de l'objet en question (commence à 1 puis est incrémentée de 1 à chaque modification).

changeset L'identifiant unique de la liste des modifications subies par l'objet.

En plus de ces 6 attributs, les 3 éléments primitifs (noeud, ligne, relation) peuvent être caractérisés par un ou plusieurs *tags*. Nous reviendrons en détail sur cet aspect dans la section 2.6. Le modèle de données présenté ici est disponible dans son intégralité à l'adresse <https://git.openstreetmap.org/rails.git/blob/HEAD:/db/structure.sql>. Un exemple d'implémentation de ce modèle est illustré dans la figure 2.3³⁷.

Pour accéder aux informations de la base de données OSM en lecture et écriture, une interface de programmation a été élaborée : l'API OSM³⁸. Il s'agit d'une API *HyperText Transfer Protocol* (HTTP) de type REST³⁹ qui en est aujourd'hui à la version 0.6⁴⁰. Plusieurs serveurs implémentent cette interface, le plus utilisé étant celui situé à l'adresse <http://www.openstreetmap.org/>. Le point d'entrée principal implémentant l'API OSM est disponible à l'adresse <http://api.openstreetmap.org/api/0.6/>. Par exemple, la requête HTTP suivante permet de retrouver tous les objets enregistrés dans la base de données OSM qui sont situés dans la boîte

34. <http://taginfo.openstreetmap.org/keys/type#values>

35. http://wiki.openstreetmap.org/wiki/Types_of_relation

36. <https://www.w3.org/TR/NOTE-datetime>

37. <http://goblor.de/wp/2009/10/16/>

38. <http://wiki.openstreetmap.org/wiki/API>

39. <https://www.w3.org/2001/sw/wiki/REST>

40. http://wiki.openstreetmap.org/wiki/API_v0.6

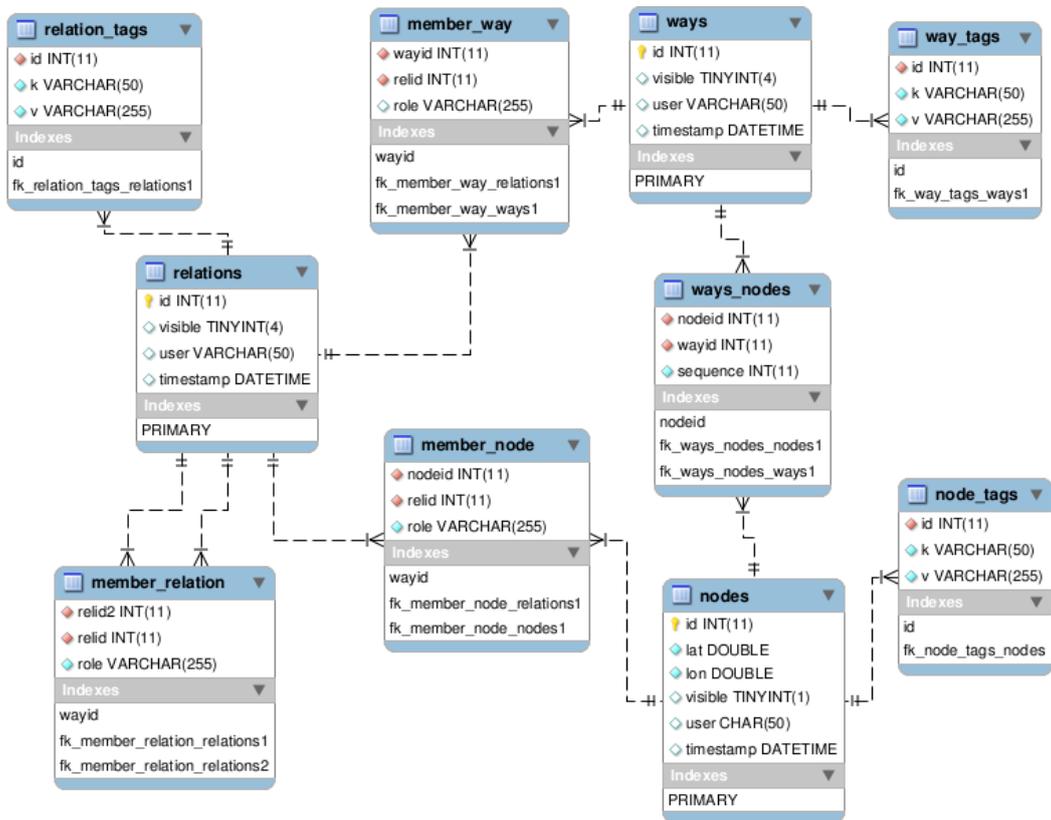


FIGURE 2.3 – Implémentation du modèle OSM avec ses 3 objets primitifs.

englobante correspondant au centre historique de la ville de Grenoble (5.723, 45.188, 5.731, 45.194) :

```
1 GET http://api.openstreetmap.org/api/0.6/map?
2   bbox=5.723,45.188,5.731,45.194
```

En retour, le serveur répond en renvoyant la liste de tous les objets demandés (en l'occurrence, 3 187 noeuds, 1 495 lignes et 248 relations pour le centre historique de Grenoble, soit une zone d'environ 0,5km²) au format XML⁴¹. Le code suivant est un extrait de cette réponse HTTP qui indique la présence d'un noeud, d'une polyligne et d'une relation entre ces derniers :

```
1 <osm version="0.6" generator="OpenStreetMap server">
2   <node id="1" changeset="1" lat="45.192" lon="5.727" version="2"
3     timestamp="2016-02-02T07:41:22Z" uid="1" user="user38">
4     <tag k="addr:housenumber" v="3" />
5   </node>
6   <way id="2" version="3" timestamp="2016-02-02T07:39:17Z"
7     changeset="2" uid="1" user="user38">
8     <nd ref="21" />
9     <nd ref="22" />
10    <nd ref="23" />
11    <tag k="access" v="permissive" />
12    <tag k="highway" v="footway" />
13    <tag k="name" v="Passage du Palais" />
14    <tag k="source" v="cadastre-dgi-fr" />
15  </way>
16  <relation id="3" version="1" timestamp="2016-02-02T07:39:17Z"
17    changeset="3" uid="1" user="user38">
18    <member type="node" ref="1" role="house" />
19    <member type="way" ref="2" role="street" />
20    <tag k="name" v="Passage du Palais" />
21    <tag k="type" v="associatedStreet" />
22  </relation>
23  ...
24 </osm>
```

Extrait de code 2.1 – Extrait de réponse en provenance d'un serveur implémentant l'API OSM au format XML.

Dans l'extrait de code 2.1, la balise `<node>` représente un noeud et les attributs `lat` et `lon` ses coordonnées géographiques. La balise `<way>` représente une polyligne dont les points sont marqués par la balise `<nd>`. Ces derniers possèdent chacun un attribut `ref` qui est la référence vers la balise `<node>` qui les représente (omise ici dans un souci de lisibilité). Enfin, la balise `<relation>` symbolise un objet de type relation. Les membres de cette relation sont référencés par la balise `<member>` dont l'attribut `ref` correspond à l'attribut `id` des objets participant à ladite relation. Cha-

41. http://wiki.openstreetmap.org/wiki/OSM_XML

cun de ces trois types d'objet possède également les attributs précédemment décrits, à savoir `id`, `uid`, `user`, `timestamp`, `version` et `changeset`. L'attribut `visible` n'est ici pas indiqué puisque, par défaut, le serveur ne récupère que les objets présents sur la dernière version de la carte OSM, la valeur de l'attribut `visible` pour tous les objets retournés est donc `true`. Pour finir, chacun des objets géographiques de cet extrait est accompagné de balises `<tag>` permettant de le caractériser, chacune d'elles composée des attributs `k` et `v` représentant respectivement sa *clé* et sa *valeur*.

2.5 Système de gestion de versions

OpenStreetMap conserve l'historique des modifications apportées à la base de données. Les modifications sont groupées dans une liste de modifications appelée *changeset*. Chacune de ces entités possède les 8 attributs suivants :

id L'identifiant unique du *changeset*.

uid L'identifiant unique de l'utilisateur à l'origine de l'ensemble des modifications.

created_at Le moment auquel les modifications ont commencé à être opérées.

closed_at Le moment auquel les modifications ont fini d'être opérées.

min_lat, min_lon, max_lat, max_lon Les quatre points formant les sommets de la boîte englobante (*Bounding Box*, BBOX) correspondant à l'emprise spatiale des modifications réalisées durant l'intervalle de temps du *changeset*.

L'API OSM évoquée dans la section 2.4 permet également de récupérer un *changeset*. La requête HTTP suivante permet de demander au serveur le 1^{er} *changeset* en base de données :

```
1 GET http://api.openstreetmap.org/api/0.6/changeset/1
```

Le serveur répond en livrant le *changeset* suivant formaté en XML :

```
1 <osm version="0.6" generator="OpenStreetMap server">
2   <changeset id="1" user="Steve" uid="1"
3     open="false" comments_count="2"
4     created_at="2005-04-09T19:54:13Z" closed_at="2005-04-09T20:54:39Z"
5     min_lat="51.528" min_lon="-0.146"
6     max_lat="51.528" max_lon="-0.146">
7   <discussion>
8     <comment date="2014-11-15T11:43:09Z" uid="2" user="John">
9       <text>From humble beginnings... <text>
10    </comment>
11    <comment date="2016-05-06T14:14:03Z" uid="3" user="Smith">
12      <text>?</text>
13    </comment>
14  </discussion>
```

```
15 </changeset>
16 </osm>
```

Extrait de code 2.2 – 1^{er} changeset de la base de données OSM.

Dans l'extrait de code 2.2, la balise `changeset` symbolise l'ensemble des modifications auquel sont associés les 8 attributs évoqués plus haut (`id`, `uid`, `created_at`, `closed_at`, `min_lat`, `min_lon`, `max_lat`, `max_lon`). Le changeset contient également un ensemble de commentaires associés (balise `<comment>`), chacun lié à un utilisateur (attribut `uid`) et à une date précise. Ces commentaires sont groupés en une discussion (balise `<discussion>`). Afin de retrouver tous les changements associés à un changeset donné, il suffit de faire la correspondance entre l'identifiant du changeset (attribut `id`) et l'attribut `changeset` d'un des 3 objets primitifs (noeud, ligne ou relation). L'API OSM permet également d'obtenir le résultat de cette jointure grâce à la requête HTTP suivante :

```
1 GET http://api.openstreetmap.org/api/0.6/changeset/1/download
```

Le serveur répond en livrant l'ensemble des modifications apportées à la base de données OSM lors du changeset dont l'identifiant est passé en paramètre de requête. La réponse est donnée au format XML `OsmChange`⁴², lequel permet d'exprimer les changements survenus sur la base de données OSM :

```
1 <osmChange version="0.6" generator="OpenStreetMap server">
2   <create>
3     <node id="1" changeset="1" timestamp="2005-04-09T19:54:13Z"
4       version="123" visible="false" user="Steve" uid="1"/>
5     <node id="2" changeset="1" timestamp="2005-04-09T19:54:39Z"
6       version="234" visible="false" user="Steve" uid="1"/>
7   </create>
8   <modify>
9     <way id="2" version="345" timestamp="2016-02-02T07:39:17Z"
10      changeset="2" uid="1" user="user38">
11       <nd ref="21" />
12       <nd ref="22" />
13       <nd ref="23" />
14       <tag k="access" v="permissive" />
15       <tag k="highway" v="footway" />
16       <tag k="name" v="Passage du Palais" />
17       <tag k="source" v="cadastre-dgi-fr" />
18     </way>
19   </modify>
20   <delete>
21     <relation id="3" version="456" timestamp="2016-02-02T07:39:17Z"
22      changeset="3" uid="1" user="user38">
23       <member type="node" ref="1" role="house" />
24       <member type="way" ref="2" role="street" />
```

42. <http://wiki.openstreetmap.org/wiki/OsmChange>

```

25 <tag k="name" v="Passage du Palais" />
26 <tag k="type" v="associatedStreet" />
27 </relation>
28 </delete>
29 </osmChange>

```

Extrait de code 2.3 – Modifications apportées à la base de données OSM lors du 1^{er} changeset.

Dans l'extrait de code 2.3, les balises `<create>`, `<modify>` et `<delete>` englobent des éléments (noeud, ligne ou relation) qui ont respectivement été ajoutés, modifiés ou supprimés en base de données.

Le projet OSM met également à disposition des fichiers de différence (*diff*) au format OsmChange⁴³. Ces fichiers regroupent les modifications effectuées sur la base de données à différentes granularités de temps (minute, heure et jour). De cette manière, une base de données OSM clone peut rester synchronisée avec les derniers changements opérés par la communauté OSM.

2.6 Les tags OpenStreetMap

A l'instar des grands médias sociaux tels que Youtube⁴⁴, Flickr⁴⁵ ou encore Tumblr⁴⁶, les plates-formes de production de données citoyennes se sont dotées de leur propre typologie de concepts. Celles-ci sont le plus souvent construites par le citoyen lui-même : on parle alors de *folksonomies*. T. Vander Wal définit une folksonomie comme suit : « le résultat d'un étiquetage (*tagging*) personnel et libre d'informations ou d'objets (tout ce qui est associé à une URL) dans le but d'une utilisation ultérieure. » [Wal, 2005]. Cet étiquetage se fait au moyen de mot-clés communément appelés des *tags*.

Comme nous l'avons vu dans la section 2.4, OSM possède un modèle de données basé sur trois composants élémentaires : le noeud (*node*), un point représenté par une longitude et une latitude ; la ligne (*way*), constituée d'un ensemble de points ; la relation (*relation*) qui est une liste ordonnée de noeuds, lignes ou d'autres relations. À chacun de ces trois éléments peuvent être associés un ou plusieurs tags permettant de le caractériser. Les tags OSM ont une forme un peu particulière : ils sont constitués de deux mots-clés - et non d'un seul comme c'est le cas sur la plupart des autres

43. <http://wiki.openstreetmap.org/wiki/Planet.osm/diffs>

44. <https://www.youtube.com/>

45. <https://www.flickr.com/>

46. <https://www.tumblr.com/>

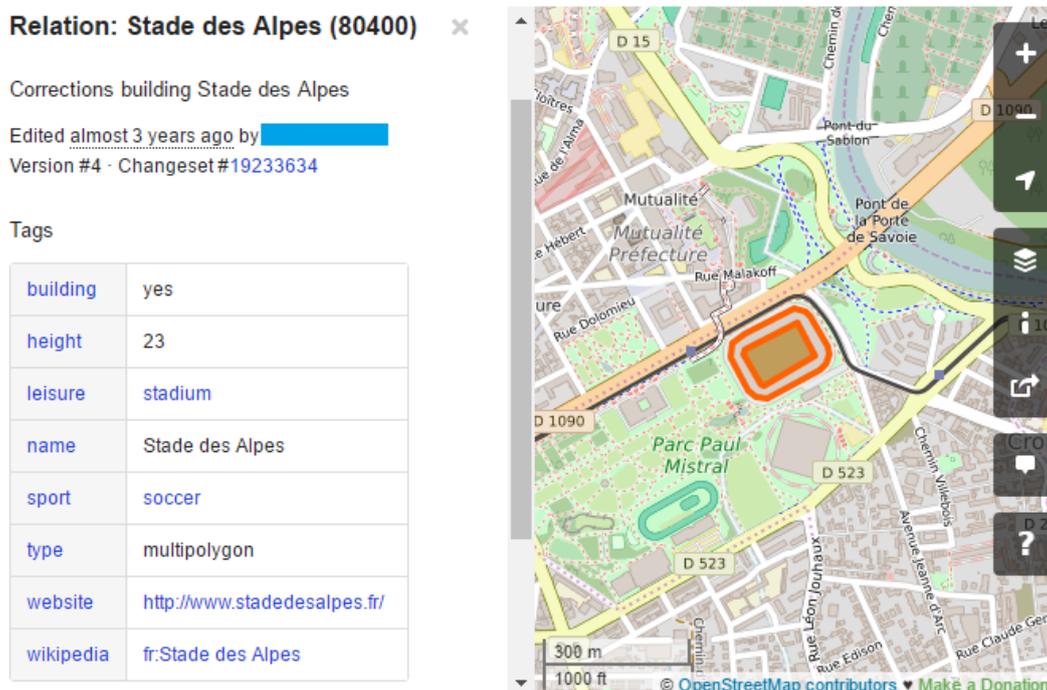


FIGURE 2.4 – Interface graphique du site Web d’OSM donnant à voir les tags d’une entité géographique (en l’occurrence le stade des Alpes de Grenoble).

médias sociaux -, formant une paire sous la forme $\langle \text{clé} \rangle = \langle \text{valeur} \rangle$. Un contributeur associera, par exemple, le tag `shop=florist` à un noeud représentant le magasin d’un fleuriste, le tag `waterway=river` à une ligne représentant un fleuve ou encore le tag `amenity=university` à une relation regroupant les différentes composantes d’un campus universitaire. La figure 2.4 est une impression d’écran de l’interface graphique du site Web d’OSM. Sur celle-ci, un objet géographique - le stade des Alpes de Grenoble - est représenté sur une carte (partie droite) accompagné de ses attributs sous la forme de tags (partie gauche).

L’ensemble de ces tags constitue la folksonomie propre à OSM. Si les ontologies qui définissent les concepts et relations du Web Sémantique sont majoritairement réalisées par les experts du domaine [Bishr and Kuhn, 2007], les folksonomies sont, quant à elles, co-construites par les internautes. Ces derniers ne les utilisent pas tant comme un moyen de faire des catégories que comme une façon de relier des éléments entre eux et de leur attribuer une signification qui fait sens à leurs yeux [Wal, 2005]. Ce système de catégorisation du contenu par les contributeurs au moyen d’étiquettes est donc totalement compatible avec la philosophie participative du projet OpenStreetMap.

La folksonomie d’OSM bénéficie d’une grande flexibilité : aussi bien la clé que la

valeur d'un tag sont définies en texte libre par les contributeurs et aucun contrôle n'est exercé au moment de la création ou la modification d'un tag dans la base de données OSM. Cette facilité d'utilisation incite les contributeurs à participer à l'enrichissement de la folksonomie. De plus, à la différence de types de classification plus stricts tels que les nomenclatures, les folksonomies sont intrinsiquement dotées d'un caractère évolutif supérieur [Shirky, 2005; ASIS&T, 2015]. En effet, l'ensemble des termes définis par une nomenclature est figé. À l'inverse, les tags constituant une folksonomie sont en perpétuel changement. Cette plasticité permet de saisir plus efficacement les transformations de la langue, elles-mêmes traduisant l'évolution de la perception du monde, et notamment l'évolution de la perception du territoire.

2.7 Limites des tags OSM

Si les folksonomies présentent d'indéniables avantages pour la classification de contenus à l'aide de tags (facilité d'utilisation, flexibilité, aspect participatif), elles souffrent, en contrepartie, de plusieurs inconvénients. Les erreurs typographiques en sont l'exemple le plus évident. En effet, aucune contrainte ne s'exerce au moment de la création ou de la modification d'un tag en texte libre. Par conséquent, la clé comme la valeur sont exposées à des fautes de frappes de la part du contributeur qui sont autant d'obstacles à une extraction des données efficace. La base de données OSM contient, par exemple, 48 tags `buiding=yes`⁴⁷ au lieu du tag `building=yes`.

Ensuite, parce qu'il n'y a pas, *a priori*, de processus permettant aux contributeurs de se mettre d'accord sur les tags à utiliser pour décrire un type d'entité, les folksonomies se heurtent au problème de la redondance conceptuelle. Pour illustrer cet autre inconvénient, prenons l'exemple des tags `leisure=swimming_pool` et `amenity=swimming_pool` qui co-existent dans la folksonomie d'OSM. Chacun d'eux décrit le même type d'objet géographique : des piscines publiques. Cependant, dans la base de données OSM, telle piscine est taggée `leisure=swimming_pool` alors que telle autre est taggée `amenity=swimming_pool`. A moins qu'il ait connaissance du recoupement entre les deux tags sus-cités, un utilisateur qui cherche à retrouver les piscines publiques sur une zone géographique donnée ne retrouvera qu'une partie de l'ensemble des piscines cartographiées s'il utilise uniquement l'un ou l'autre de ces tags.

D'autre part, l'usage des tags peut être ambigu. S. Golder identifie sept fonctions différentes des tags, chacune liée à une intention différente de l'utilisateur [Golder,

47. <http://taginfo.openstreetmap.org/tags/buiding=yes>

2006] : certains tags servent à caractériser le sujet de la ressource taggée, d'autres à reconnaître son type, d'autres encore à identifier son créateur. La réutilisation de cette information devient alors d'autant plus difficile qu'elle repose sur une incertitude quant à ce qu'a voulu exprimer le tagger. Dans le cas de la folksonomie OSM, on trouve d'autres types d'ambiguïté. Par exemple, certains tags donnent une information sur la nature de l'objet géographique auquel ils sont associés (par exemple, les tags `amenity=restaurant`, `tourism=hotel` ou encore `historic=castle`), tandis que d'autres n'expriment qu'une de ses propriétés (par exemple, les tags `name=StadedesAlpes`, `internet_access=yes` ou encore `opening_hours=9am-5pm`).

Enfin, les folksonomies ne permettent pas d'exprimer des relations entre les tags qui la composent. Pourtant, il existe bel et bien des relations *implicites* entre certains tags voire entre les composantes de ces tags (clé et valeur). Par exemple, il existe un lien de proximité entre les tags `shop=bicycle` et `amenity=bicycle_parking` puisqu'ils ont en commun d'appartenir au *domaine du cyclisme*. Par ailleurs, un lien de subsomption peut être dégagé entre la clé et la valeur du tag `shop=bicycle` : les *magasins de vélo* sont une catégorie particulière de *magasins*. D'autre part, le tag `cuisine=chinese` n'a de sens que s'il est combiné avec le tag `amenity=restaurant` dont il est dépendant. L'élaboration d'un modèle qui décrive *explicitement* ces relations sémantiques faciliterait à la fois le contrôle de la qualité des données au moment de l'acquisition, mais permettrait également une interrogation plus efficace de la base de données OSM.

Malgré ces inconvénients, se priver de ce type de classification serait nettement préjudiciable puisque la description de l'information perdrait, dès lors, son caractère évolutif [Ho and Rajabifard, 2010]. Afin de tirer le meilleur parti de la folksonomie d'OSM, nombre de travaux ont élaboré des méthodes et modèles qui s'efforcent de pallier les désavantages liés à la catégorisation de contenu par tags [Auer et al., 2009; Codescu et al., 2011; Baglatzi et al., 2012; Ballatore et al., 2013]. Un guide de bonnes pratiques sous la forme d'un site Web de type Wiki est maintenu par les contributeurs OSM afin de tenir à jour des tags dont l'usage a fait l'objet d'un consensus (et éventuellement d'une discussion préalable). Les travaux de S. Auer [Auer et al., 2009], M. Codescu [Codescu et al., 2011] et A. Ballatore [Ballatore et al., 2013] tentent d'augmenter l'expressivité formelle de la folksonomie OSM en ajoutant certaines relations (subsomption, similarité) entre les tags et leur composantes (la clé et la valeur) pour en construire une ontologie. A. Baglatzi, dans [Baglatzi et al., 2012], s'efforce de rattacher les concepts exprimés par le biais des tags OSM à une ontologie de haut niveau au moyen d'un questionnaire demandant au contributeur qui utilise un tag d'explicitement sa sémantique. Ces travaux présentent des avancées significatives pour la représentation des connaissances dans l'Information Géographique

que nous analysons en détail dans le chapitre 3. Ils présentent également certaines limites dont nous définissons l'emprise dans le chapitre 4. Nos travaux s'inscrivent dans la lignée de ces efforts en proposant un méta-modèle permettant de saisir les spécificités des tags OSM, et en proposant de construire une ontologie à partir de ces tags qui dépasse les limites des travaux sus-cités.

2.8 Conclusion

Dans ce chapitre, nous avons, dans un premier temps, présenté un historique (section 2.2) du projet de cartographie participative OpenStreetMap (section 2.2) depuis sa création en 2004 jusqu'au succès qu'il connaît aujourd'hui dans le domaine de la VGI. Ensuite, ses principales caractéristiques ont été exposées : échelle infra-communale des données, expertise terrain de ses contributeurs, collecte de l'information en continu et cadre juridique assurant un accès gratuit et ouvert à toute l'information qui en est issue. D'autre part, nous avons détaillé le modèle qui soutient la base de données OSM avec ses 3 types primitifs : le noeud, la ligne et la relation (section 2.4). Puis le système de gestion des versions de cette même base a été expliqué (section 2.5).

Enfin, les principes de la catégorisation de contenu par tags tels qu'ils sont mis en œuvre dans le projet OSM ont été présentés. La dernière partie de ce chapitre aborde les limites à l'utilisation de folksonomies : erreurs typographiques, redondance conceptuelle, ambigüité sémantique et manque d'expressivité formelle. Dans le chapitre suivant, nous présentons l'état de l'art des travaux effectués pour pallier ces inconvénients.

Chapitre 3

Méthodes de structuration de la folksonomie OpenStreetMap

3.1 Introduction

Dans le chapitre précédent, nous avons exposé les principes du projet de cartographie participative OpenStreetMap. Dans un premier temps, nous avons présenté une brève rétrospective de son évolution depuis ses débuts en 2004 jusqu'au succès qu'il connaît aujourd'hui - désormais véritable fer de lance de l'Information Géographique Volontaire. Dans un second temps, nous avons détaillé le modèle qui structure sa base de données et, tout particulièrement, sa façon de catégoriser les objets géographiques collectés par le biais d'un ensemble de tags : la folksonomie OSM.

Plusieurs travaux ont abordé les problèmes liés à la folksonomie d'OpenStreetMap : erreurs typographiques, redondance, ambiguïté, manque d'expressivité formelle. Le présent chapitre est consacré à l'état de l'art de ces différents efforts de sémantisation des tags OSM. En premier lieu, le guide de bonnes pratiques accessible sur le wiki OSM est présenté (section 3.2). Ensuite, les ontologies *Linked GeoData* (section 3.3) et OSMonto (section 3.4) sont expliquées. Puis, nous exposons les principes d'une méthode d'alignement semi-automatique sur une ontologie haut-niveau au moyen d'un jeu avec un but (section 3.5). Enfin, ce chapitre s'achève sur la section 3.6 avec la présentation de l'ontologie OSN qui constitue le modèle sémantique basé sur les tags OSM le plus abouti.

3.2 Wiki OpenStreetMap

Afin de décrire aussi précisément que possible les caractéristiques physiques du terrain, les contributeurs d'OSM doivent partager un vocabulaire commun. Ce vocabulaire est exprimé au moyen d'une paire clé-valeur de mots-clés aussi appelée tag. Ces tags sont formulés en texte libre et aucune contrainte n'est exercée lors de l'ajout, de la modification ou de la suppression d'un tag en base de données. Cette méthode garantit une grande flexibilité, mais est également un obstacle majeur à l'exploitation de la folksonomie OSM (cf. section 2.6). Dans l'optique de dresser un inventaire de ces tags et de leur usage, le projet OSM s'appuie sur un guide de bonnes pratiques sous la forme d'un wiki¹. La figure 3.1 est un extrait de la page *Map features*² qui liste les tags référencés sur le wiki OSM. En l'occurrence, il s'agit de tags ayant trait à l'éducation.

Clé	Valeur	Education			
amenity	college		A college campus or buildings		
amenity	kindergarten		For children too young for a regular school (also known as playschool or nursery school).		
amenity	library		A public library (municipal, university, ...) to borrow books from.		
amenity	school		School and grounds		
amenity	university		A university campus		

FIGURE 3.1 – Extrait du guide de bonnes pratiques pour l'utilisation des tags sur le Wiki OSM.

D'après [Ballatore et al., 2013], en 2013, ce wiki regroupait 2047 tags, sur les 77 millions de tags³ existants. Pour chaque tag répertorié sur la page *Map Features*, les 6 caractéristiques suivantes sont données :

Clé La clé du tag.

Valeur(s)/Type de valeur La valeur du tag. Dans le cas où il n'existe qu'un nombre de valeurs très restreint pour une clé donnée, toutes les valeurs sont

1. <http://wiki.openstreetmap.org/>
2. http://wiki.openstreetmap.org/wiki/Map_Features
3. <http://wiki.openstreetmap.org/wiki/Taginfo/FAQ>

listées. Si le panel de valeurs est très large, c'est le type qui est indiqué (date, numérique, booléen).

Type(s) d'objet associé(s) Un ou plusieurs type(s) d'objet OSM au(x)quel(s) les tags sont susceptibles d'être associés : noeud, ligne et/ou relation.

Définition et usage Définition du type d'objet géographique qui doit être caractérisé par le tag. Des précisions peuvent être apportées sur l'usage du tag, en particulier sur les tags avec lesquels le tag décrit peut être combiné.

Rendu Le rendu graphique préconisé sur une carte associé à un objet possédant le tag.

Photo Illustration du tag par la photo d'un exemple du type d'objet qu'il caractérise.

Cette formalisation de l'usage des tags permet de désambiguïser leur usage et, *a fortiori*, de faciliter l'exploitation de la base de données.

La table des matières de la page *Map Features* sous laquelle sont rangés les différents tags est hiérarchisée en trois niveaux. Le premier divise les tags en deux classes : les tags de type "Caractéristique principale" ("*Primary Feature*") et les tags de type "Propriété complémentaire" ("*Additional property*"). Le premier type de tag donne une information sur la nature même de l'objet auquel il est associé. Par exemple, `amenity=library` pour décrire une bibliothèque, `shop=bakery` pour caractériser une boulangerie ou encore `craft=carpenter` pour indiquer un charpentier. Le second type regroupe des tags qui permettent de préciser certaines propriétés de l'objet cartographié qui s'ajoutent à au moins un tag de type "Caractéristique principale". Par exemple, les tags `name=Grenoble`, `wikipedia=https://en.wikipedia.org/wiki/Grenoble` et `population=162780` permettent respectivement d'indiquer que l'élément auquel ils sont associés est nommé "Grenoble", possède une page sur l'encyclopédie Wikipédia à l'URL indiquée et compte une population de 162 780 habitants (cet élément possède également un tag de type "Caractéristique principale" `place=city`).

Les deux et troisième sous-niveaux de la catégorie "Caractéristique principale" correspondent respectivement à des thèmes et sous-thèmes. Le deuxième niveau est composé de catégories telles que *Amenity*, *Craft* ou encore *Leisure*. Le troisième niveau dispose de sous-thèmes plus précis que le second niveau. Ainsi, sous la catégorie *Primary Feature* > *Amenity*, se trouvent les sous-catégories *Sustenance*, *Education*, *Transportation*, *Financial*, *Healthcare*, *Entertainment* et *Arts & Culture*. Sous la catégorie *Primary Feature* > *Amenity* > *Education*, se trouvent, par exemple, les tags `amenity=kindergarten`, `amenity=school` ou encore `amenity=college`. Les clés des tags correspondent au deuxième niveau de la table des matières de la page de docu-

mentation des tags OSM. Les premier et troisième niveaux de cette classification ne correspondent *a priori* à rien dans la forme des tags. Finalement, cet arrangement des tags permet aux contributeurs de standardiser de façon rudimentaire l'utilisation des tags OSM afin d'éviter, dans une certaine mesure, la redondance inhérente aux folksonomies, frein majeur à l'extraction des données. Le Wiki OSM permet également de parcourir les tags de manière thématique et, par conséquent, d'affiner la caractérisation des objets géographiques au moment de l'édition de l'information en base de données.

Comme n'importe quel wiki, le site Web de documentation OSM supporte la méthode de discussion de contenu par les contributeurs : le cycle Contribution/Révo-cation/Discussion (*Bold/Revert/Discuss*, BRD)⁴. Ainsi, les directives sur l'usage des tags OSM sont établies par les contributeurs eux-mêmes, et chaque tag choisi pour y figurer peut être sujet à discussion dans le but d'aboutir à un consensus. Il s'agit ici de parvenir à un "accord social" nécessaire pour définir des symboles sémantiques communs pouvant être compris par la plupart des utilisateurs [Kuhn, 2005] et désamorcer d'éventuelles "guerre de tags" (*tag wars*) selon les termes de P. Mooney et P. Corcoran dans [Mooney and Corcoran, 2012]. Le cycle BRD permet aux utilisateurs impliqués dans un contentieux de créer une page de discussion associée au tag concerné afin d'argumenter leur choix et d'arriver, *in fine*, à un accord.

Plus spécifiquement, le wiki OSM préconise de suivre un processus de proposition de tag détaillé⁵. Ce dernier consiste en plusieurs étapes représentées par un statut⁶ (en gras dans la liste ci-dessous) associé à la proposition :

1. Vérifier que le tag n'existe pas déjà sur le Wiki.
2. **Draft** Créer une page de proposition du tag à l'aide du modèle disponible sur le Wiki⁷.
3. **Proposed** Soumettre la proposition sur la liste de diffusion prévue à cet effet⁸, puis envoyer une demande de commentaires (*Request For Comments*, RFC). Discuter le bien fondé de la proposition avec la communauté pendant deux semaines.
4. **Voting** Procéder au vote de la proposition pendant deux semaines reconduc-tibles en cas de manque de réponses. Les raisons d'un vote contre doivent être argumentées.

4. https://en.wikipedia.org/wiki/Wikipedia:BOLD,_revert,_discuss_cycle

5. http://wiki.openstreetmap.org/wiki/Proposal_process

6. http://wiki.openstreetmap.org/wiki/Category:Proposals_by_status

7. http://wiki.openstreetmap.org/wiki/Template:Proposal_Page

8. <https://lists.openstreetmap.org/listinfo/tagging>

5. **Post-vote** Dans le cas où le dépouillement du vote et la prise en compte des commentaires ne sont pas immédiats, il convient de changer le statut de la proposition de `voting` vers `post-vote`.
6. **Approved** Dépouiller le vote et prendre en compte les éventuels commentaires. Modifier le statut : `approved` si le tag est accepté, `rejected` sinon.
7. **Abandoned** Si la procédure de proposition de tag s'éternise (plus de 3 mois), elle doit être abandonnée.
8. **Canceled** Si le créateur souhaite retirer sa proposition en cours, il peut l'annuler.
9. **Obsolete** Si une autre proposition rend la proposition courante vaine, elle doit être annotée comme obsolète.
10. **Undefined** Lorsque le statut d'une proposition de tag n'a pas été renseigné par son auteur, il faut l'indiquer en déclarant son statut inconnu.

Ainsi, l'ajout d'un tag litigieux par un utilisateur peut potentiellement être annulé ou modifié si l'ensemble des contributeurs en convient. Cette méthode répond parfaitement aux exigences de l'approche ascendante promue par le projet OSM.

Finalement, pour chaque clé comme pour chaque tag inventorié(e) sur la page *Map Features* du wiki OSM, il existe une page dédiée sur le wiki OSM. Les URL de ces pages sont formalisées : pour retrouver une clé, il convient d'utiliser l'URL de la forme `http://wiki.openstreetmap.org/wiki/Key:<clé>`, tandis que pour retrouver des informations sur un tag, on utilisera une URL de la forme `http://wiki.openstreetmap.org/wiki/<tag>`. Sur chacune de ces pages, la description du type d'objet à associer avec la clé ou le tag ainsi que des précisions sur l'usage de celui-ci sont données. De plus, des données structurées décrivent certains attributs de la clé ou du tag au moyen d'une *infobox* telle que celles que l'on retrouve dans la plupart des sites de type Wiki. Cette infobox donne les informations suivantes :

Description La description du type d'objet auquel doit être associé le tag.

Type(s) d'objet associé(s) Un ou plusieurs type(s) d'objet OSM au(x)quel(s) les tags sont susceptibles d'être associés : noeud, ligne et/ou relation.

Nombre de valeurs documentées (ne concerne que les clés) Nombre de valeurs documentées associées à la clé.

Combinaisons utiles Les combinaisons préconisées avec d'autres tags.

Statut Statut relatif à l'acceptation du tag par la communauté OSM.

Statistiques d'utilisation Un aperçu des statistiques d'utilisation fournies par le service Web TagInfo⁹.

9. <http://taginfo.openstreetmap.org/>

Les informations disponibles sur le wiki OSM centralisent et formalisent les règles d'usage des tags OSM, et ce de façon participative. Ce guide de bonnes pratiques permet à l'utilisateur consciencieux d'homogénéiser les méta-données sur les objets géographiques cartographiés, ouvrant la voie à une exploitation plus efficace de la base de données OSM.

3.3 *Linked GeoData* (LGD)

Les travaux des auteurs de [Auer et al., 2009] au sein du groupe de recherche pour le Web Sémantique et l'Ingénierie de la Connaissance Agile (*Agile Knowledge Engineering and Semantic Web research group*, AKSW)¹⁰ comptent parmi les premières tentatives de construction d'une ontologie à partir de l'information en base de données OSM. L'ontologie *Linked GeoData* (LGD)¹¹ en est l'aboutissement dont la première version est publiée en 2009. Dans ce modèle, les tags sont classés en trois catégories, chacune correspondant à un modèle de conversion vers le langage de représentation des connaissances *Ontology Web Language* (OWL)¹² :

Attributs de classification Ce sont des tags qui donnent des informations sur la nature des éléments spatiaux auxquels ils sont associés (par exemple, `amenity=school`). Ce sont les tags regroupés sous la catégorie "Caractéristique principale" sur la page *Map Feature* du wiki OSM. Dans l'ontologie LGD, la clé et la valeur sont toutes deux représentées par des classes (`owl:Class`). La clé est considérée comme super-classe de la valeur à laquelle elle est associée dans la paire `clé=valeur`. La relation de subsumption est décrite au moyen de la propriété `rdfs:subClassOf` ; ainsi les tags `amenity=school` et `amenity=kindergarten` vont donner lieu à une hiérarchie à deux niveaux composée d'un concept (`owl:class`) `Amenity` super-classe des concepts `School` et `Kindergarten`.

Attributs de description Ce sont des tags ayant pour valeur un ensemble de valeurs prédéfinies (par exemple, `internet_access=wired/wlan/terminal`). Ce sont des tags regroupés sous la catégorie "Propriété complémentaire" sur la page *Map Feature* du wiki OSM. Ils sont convertis en propriétés objets (`owl:ObjectProperty`) dans l'ontologie LGD ;

Attributs de données Ce sont des tags dont la valeur est textuelle ou de type primitif (par exemple, `height=80`). Ces tags sont eux aussi rangés sous la ca-

10. <http://aksw.org/>

11. <http://wiki.openstreetmap.org/wiki/LinkedGeoData>

12. <http://www.w3.org/2001/sw/wiki/OWL>

tégorie "Propriété complémentaire" sur la page *Map Feature* du wiki OSM. Ils sont convertis en propriétés primitives (`owl:DataProperty`) dans l'ontologie.

Plus qu'un méta-modèle, l'ontologie LGD est également une vaste base de données géographiques. En effet, ce ne sont pas uniquement les concepts (la boîte de terminologie (*Terminology Box*, TBox), c'est-à-dire les tags OSM) qui sont représentés : les instances (la boîte d'assertion (*Assertion Box*, ABox), c'est-à-dire les objets géographiques) font également partie intégrante de l'ontologie. Ainsi, chaque élément primitif de la base de données OSM (noeud, ligne, relation) est décrit par autant de triplets RDF qu'il a de tags associés. Par conséquent, chaque objet est défini dans le *triplestore* comme une instance d'un ou plusieurs attribut(s) de classification et ayant un ou plusieurs attribut(s) de description et de données.

De plus, les auteurs de [Auer et al., 2009] ont développé une heuristique qui tente d'établir une relation d'équivalence (`owl:sameAs`) entre chaque instance de l'ontologie LGD et l'entité de la base de données DBpedia¹³ à laquelle elle correspond. DBpedia est une base de connaissances qui vise à structurer l'information extraite depuis l'encyclopédie libre Wikipédia¹⁴. Les informations qui sont formalisées apparaissent ensuite dans les articles Wikipédia sous la forme d'une *infobox*¹⁵ dans la partie droite de l'interface qui synthétise les caractéristiques essentielles du sujet traité dans l'article. Le but de la structuration proposée par le projet DBpedia est de permettre une interrogation sophistiquée des données massives regroupées dans l'encyclopédie Wikipédia.

La mesure de similarité entre un objet de l'ontologie LGD et une entité provenant de DBpedia est calculée sur la base de la comparaison de leur nom, de leur localisation et de leur type. De cette manière, plus de 50 000 objets de l'ontologie LGD sont identifiés sur une base de données largement reconnues dans l'écosystème des Données Ouvertes et Liées. Prise dans son ensemble, l'ontologie LGD contient 500 classes (issues de tags de type attribut de classification), 50 propriétés objets (issues de tags de type attribut de description) et 15 000 propriétés primitives (issues de tags de type attribut de données).

Pour illustrer cette méthode de conversion des données OSM au format RDF, considérons l'exemple donné par l'extrait de code 3.1 représentant l'église Notre-Dame de Dresde telle qu'elle est retournée par un serveur implémentant l'API OSM au format XML :

13. <http://dbpedia.org/>

14. <http://wiki.dbpedia.org/about>

15. <https://en.wikipedia.org/wiki/Help:Infobox>

```

1 <node id="123" lat="51.051" lon="13.741" version="10" changeset="123"
2   user="John" uid="123" visible="true" timestamp="2009-03-09T08:49:48Z">
3   <tag k="name" v="Frauenkirche" />
4   <tag k="created_by" v="Potlatch 0.10e" />
5   <tag k="tourism" v="viewpoint" />
6   <tag k="url" v="http://www.frauenkirche-dresden.de/" />
7   <tag k="denomination" v="lutheran" />
8   <tag k="wikipedia:en" v="Frauenkirche_Dresden" />
9   <tag k="religion" v="christian" />
10  <tag k="amenity" v="place_of_worship" />
11  <tag k="wikipedia:de" v="Frauenkirche_(Dresden)" />
12 </node>

```

Extrait de code 3.1 – Représentation de l’église Notre-Dame de Dresde au format XML par un serveur implémentant l’API OSM.

L’extrait de code 3.2 ci-dessous est le résultat de la conversion du noeud correspondant à l’identifiant 123, c’est-à-dire à la représentation de l’église Notre-Dame de Dresde telle que décrite dans la base de données OSM (cf. extrait de code 3.1), au format RDF dans le triplstore de l’ontologie LGD :

```

1 lgd-node:123 rdfs:comment "Generated by Triplify V0.5" .
2 lgd-node:123 cc:license cc:by-sa/2.0 .
3 lgd-node:123 lgd-vocabulary:attribution "This data is derived" .
4 lgd-node:123#id rdf:type lgd-vocabulary:node .
5 lgd-node:123#id geo-wgs84:long "13.7416"^^xsd:decimal .
6 lgd-node:123#id geo-wgs84:lat "51.0519"^^xsd:decimal .
7 lgd-node:123#id lgd-vocabulary:created_by lgd:Potlatch+0.10e .
8 lgd-node:123#id lgd-vocabulary:religion lgd:christian .
9 lgd-node:123#id lgd-vocabulary:name "Frauenkirche" .
10 lgd-node:123#id lgd-vocabulary:tourism lgd:viewpoint .
11 lgd-node:123#id lgd-vocabulary:amenity lgd:place_of_worship .
12 lgd-node:123#id lgd-vocabulary:wikipedia%2525de
13   "http://de.wikipedia.org/wiki/Frauenkirche_(Dresden)" .
14 lgd-node:123#id lgd-vocabulary:wikipedia%2525en
15   "http://en.wikipedia.org/wiki/Frauenkirche_Dresden" .
16 lgd-node:123#id lgd-vocabulary:denomination lgd:lutheran .
17 lgd-node:123#id lgd-vocabulary:url "http://www.frauenkirche-dresden.de/" .
18 lgd-node:123#id lgd-vocabulary:locatedNear lgd-way:23040893> .
19 lgd-node:123#id lgd-vocabulary:locatedNear lgd-way:23040894> .

```

Extrait de code 3.2 – Triplets représentant l’église Notre-Dame de Dresde dans l’ontologie LGD.

Pour finir, l’ontologie LGD constitue une base de plus de 1,2 milliards de triplets dans sa dernière version¹⁶, libre de droit et disponible gratuitement sous licence ODbL¹⁷. La mise à jour des données est faite tous les ans ou tous les deux ans. De plus, l’information est directement accessible via un point d’entrée pour le

16. <http://blog.aksw.org/linkedgeodata-new-rdf-versions/>

17. <http://downloads.linkedgeodata.org/>

langage d’interrogation de données au format RDF SPARQL¹⁸. Enfin, l’ontologie LGD dispose d’une interface graphique *ad hoc* pour l’exploration et la modification de ces données¹⁹, permettant au profane de bénéficier de l’expressivité formelle de l’ontologie, notamment grâce à un système de suggestion de tags. En effet, lorsque l’utilisateur veut associer le tag `amenity=school` à une école maternelle, le système est capable de repérer que ce tag est lié à l’attribut de classification `School`, sous-classe de `Amenity`, et peut lui proposer des tags similaires, comme d’autres tags liés à des sous-concepts du concept `Amenity`, peut-être plus pertinents, tels que `amenity=kindergarten`. Cette approche contribue à l’amélioration de la qualité des méta-données de la base de données OSM.

3.4 OSMonto

L’ontologie OSMonto²⁰ [Codescu et al., 2011] a été créée en 2011 par l’Université de Bremen²¹ et le Centre de Recherche Allemand pour l’Intelligence Artificielle (DFKI)²². Elle a pour but d’appuyer un outil de navigation centré sur les activités proposées par les objets géographiques cartographiés dans OpenStreetMap²³. Avec cet outil, l’utilisateur qui souhaite aller d’un point A à un point B indique, dans un premier temps, 1) son point de départ, 2) sa destination et 3) les activités auxquelles il souhaite s’adonner (restauration, sport, activités culturelles, *etc.*). Alors, le système propose le meilleur itinéraire en fonction des préférences renseignées par l’utilisateur, des activités disponibles selon la base de données OSM, et de la distance de ces activités par rapport au trajet à parcourir.

Dans cette ontologie, seuls les tags qui possèdent plus de 100 occurrences dans la base de données OSM, ainsi que ceux qui sont référencés dans le wiki OSM, sont intégrés dans l’ontologie. Ce parti pris assure une meilleure qualité des tags considérés puisqu’ils sont *de facto* plus consensuels et qu’«il n’y a jamais 100 fois la même erreur» [Codescu et al., 2011]. Cela permet, par exemple, d’écarter les tags redondants qui seraient le fruit d’une méconnaissance des bonnes pratiques OSM, ou encore des tags contenant des erreurs typographiques.

La méthodologie qui sous-tend sa création est semblable à celle de l’ontologie LGD en cela que les concepts extraits des clés des tags OSM sont considérés comme

18. <https://fr.wikipedia.org/wiki/SPARQL>

19. <http://browser.linkedgeo.org/>

20. <http://wiki.openstreetmap.org/wiki/OSMonto>

21. <http://www.uni-bremen.de/>

22. <http://www.dfki-bremen.de/>

23. <http://do-roam.org/>



FIGURE 3.2 – Aperçu de la hiérarchie de tags issue de l’ontologie OSMonto.

super-classes des concepts extraits des valeurs associées. Ainsi, à partir des tags `shop=supermarket` et `shop=convenience`, la classe `k_shop` est créée, super-classe des classes `v_supermarket` et `v_convenience`. Cette approche a été adoptée chaque fois que la valeur du tag était une constante OSM plutôt qu’une chaîne de caractères ou un nombre. Ainsi, les tags OSM classifiés comme attributs de description ou attributs de données ne sont pas pris en considération dans la méthodologie de construction de l’ontologie OSMonto.

Afin de prévenir l’ambiguïté dans le cas où elles sont les mêmes, les classes (`owl:class`) issues des clés ou bien des valeurs sont respectivement préfixées par `k_` et `v_`. Ainsi, le tag utilisé pour décrire, par exemple, une gare sur une ligne ferroviaire au sens large, `railway=station`, donne la classe `v_station` tandis que le tag employé pour spécifier, par exemple, le type de gare dont il s’agit, par exemple `station=subway` pour un arrêt de métro, donne la classe `k_station`.

Les dépendances entre tags sont également considérées : certains tags ne sont compatibles qu'avec d'autres tags bien spécifiques. Par exemple, le tag caractérisant un type de cuisine `cuisine=*` (par exemple, `cuisine=seafood`), n'a de sens qu'au sein d'une combinaison avec le tag `amenity=restaurant`. Dans ces cas, la clé du tag dépendant est convertie en une propriété objet `owl:ObjectProperty` ayant pour domaine la valeur du tag dont il est dépendant et sa propre valeur pour rang. Par conséquent, pour exprimer la cooccurrence des tags `amenity=restaurant` et `cuisine=*`, la propriété objet `hasCuisine` est créée avec pour domaine la classe `v_restaurant` et pour image la classe `k_cuisine`. De cette manière, et puisque toutes les valeurs associées à la clé `cuisine` (`french`, `seafood`, `pizza`) sont des sous-classes de la classe `k_cuisine`, elles peuvent être utilisées comme valeur de la propriété `hasCuisine`.

Pour finir, l'ontologie OSMonto est utile à plusieurs égards. Tout d'abord, elle donne un aperçu général des tags caractérisant la nature des objets géographiques de la base OSM sous la forme d'une hiérarchie à deux niveaux clé-valeur illustrée en figure 3.2. D'autre part, les auteurs tirent parti du langage OWL pour unifier des concepts similaires exprimés par différents tags OSM. Ils établissent, par exemple, une relation d'équivalence `owl:equivalentTo` entre les classes extraites des tags `fuel:electricity=yes` (associé à des stations essence également équipées de bornes pour voitures électriques) et `amenity=charging_station` (associé à des bornes de rechargement de voitures électriques), tous deux associés à des équipements permettant de recharger des voitures électriques. Cette technique permet de faire le lien entre des classes entre lesquelles aucune similarité ne peut *a priori* être inférée à partir de la hiérarchie clé-valeur puisque elles sont issues de tags ne partageant pas la même clé.

Ces correspondances peuvent être établies de manière automatique grâce à des *matchers* d'ontologies. Le matcher Falcon²⁴ a, par exemple, permis de retrouver 80% des relations équivalences entre les classes de OSMonto jugées pertinentes par les auteurs de [Codescu et al., 2011]. Toutefois, ces outils ont leurs limites : à cause d'une distance orthographique trop importante, la correspondance entre les tags `fuel:electricity=yes` et `amenity=charging_station` n'est pas retournée par Falcon. Aussi, les auteurs préconisent une méthode semi-automatique dans laquelle les *matchers* assistent les utilisateurs dans leur prise de décision. Par le biais de cette technique d'alignement, les auteurs proposent également d'établir des relations d'équivalence entre les classes de OSMonto et d'autres ontologies du LOD, et

24. <http://ws.nju.edu.cn/falcon-ao/>

notamment des ontologies de haut niveau telle que OpenCyc²⁵, GUM-Space²⁶ ou encore Dolce²⁷. La construction de ponts entre ces différentes représentations de la connaissance faciliterait l'interopérabilité des tags OSM dans le Web Sémantique.

3.5 Game With A Purpose (GWAP)

Comme suggéré par les créateurs de l'ontologie OSMonto dans [Codescu et al., 2011], les auteurs de [Baglatzi et al., 2012] proposent une méthode d'alignement des tags OSM sur l'ontologie de haut niveau Dolce+DnS Ultra Lite (DUL)²⁸. Une ontologie générique (de haut niveau) permet notamment de faire le pont entre différentes ontologies d'un même domaine développées indépendamment et dont les vocabulaires ne sont pas immédiatement compatibles [Du et al., 2013]. Le but de cet effort est, donc, de rendre les tags OSM interopérables avec les ontologies du Web Sémantique.

L'ontologie DUL qui est ici mise à contribution repose sur deux ontologies de haut niveau. La première, *Descriptive Ontology for Linguistic and Cognitive Engineering* (DOLCE)²⁹, a été développée en 2002 par les auteurs de [Gangemi et al., 2002] du Laboratoire pour les Ontologies Appliquées (LOA)³⁰ dans le cadre du projet WonderWeb³¹. Ce projet vise à créer une bibliothèque d'ontologies de concepts fondamentaux représentant des points de vue ontologiques différents. Le but est d'interopéabiliser les ontologies de domaine du Web Sémantique tout en conservant la possibilité d'interpréter les concepts selon divers biais primitifs explicités dans chacune de ces ontologies de haut niveau.

L'ontologie DOLCE présente un biais cognitif clair puisque les concepts et relations qui la constituent représentent des catégories ontologiques exprimées dans le langage humain et dans le sens commun. Ainsi, les concepts de cette ontologie n'expriment pas les qualités intrinsèques du monde (vision prescriptive) mais bien des "artefacts cognitifs issus de de la perception humaine, d'empreintes culturelles ou encore de conventions sociales" (vision descriptive)³².

25. <http://sw.opencyc.org/>

26. <http://ontologydesignpatterns.org/wiki/Ontology:GUM-Space>

27. <http://www.loa.istc.cnr.it/old/DOLCE.html>

28. <http://www.ontologydesignpatterns.org/ont/dul/DUL.owl>

29. <http://www.loa.istc.cnr.it/old/DOLCE.html>

30. <http://www.loa.istc.cnr.it/>

31. <http://wonderweb.man.ac.uk/>

32. https://en.wikipedia.org/wiki/Descriptive_Ontology_for_Linguistic_and_Cognitive_Engineering

La seconde ontologie sur laquelle est bâtie DUL, l'*Descriptions and Situations ontology* (DnS)³³ est en réalité une extension de l'ontologie DOLCE. DnS adopte un point de vue constructiviste en cela que les axiomes exprimés selon son méta-modèle sont associés à une situation, insistant d'autant plus sur la notion de point de vue lié à un contexte spécifique [Gangemi et al., 2002; Gangemi, 2009].

Afin d'aligner les tags OSM sur la super ontologie DUL, les auteurs de [Baglatzi et al., 2012] introduisent une méthode pour évaluer le degré de similarité entre un tag et un concept de l'ontologie DUL via un jeu avec un but (*Game With A Purpose*, GWAP). Celui-ci consiste à poser une série de questions sur la nature d'un objet géographique ajouté par un contributeur dans la base de données OSM. Chacune de ces questions correspond en réalité à un concept de l'ontologie DUL. A titre d'exemple, la question "S'agit-il d'un objet physique tel qu'une rivière ou un stade?" fait référence au concept `dul:PhysicalObject`. Cette technique dissimule la complexité de l'ontologie à l'utilisateur et lui permet de participer à la description des tags OSM compatibles avec les schémas fondamentaux sur lesquels reposent le Web Sémantique.

3.6 *OSM Semantic Network* (OSN)

Enfin, le réseau sémantique OSN³⁴ élaboré par les auteurs de [Ballatore et al., 2013] compte parmi les travaux les plus récents pour la structuration de la folksonomie OSM. L'ontologie est exprimée dans le langage de représentation des connaissances *Simple Knowledge Organization System* (SKOS).

Les auteurs de cette approche ont développé un robot d'indexation (*Web crawler*) qui parcourt le wiki OSM et explore chacune de ses pages. Ces dernières sont groupées en 5 catégories :

Page clé Ce type de page est dédié à la signification et à l'usage des clés des tags.

Page tag Ce type de page est consacré à la signification et à l'usage d'un tag.

Page proposition de tag Ce type de page est consacré à la signification et à l'usage d'un tag proposé à la communauté OSM et contient les arguments soulevés au cours du processus de délibération quant à l'acceptation ou au rejet dudit tag.

Page cluster Ce type de page correspond au deuxième niveau dans la hiérarchie décrite par la table des matières de la page *Map Features*³⁵ du wiki OSM,

33. <http://www.loa.istc.cnr.it/ontologies/ExtendedDnS.owl>

34. http://spatial.ucd.ie/osn/skos/osm_semantic_network.skos.rdf

35. http://wiki.openstreetmap.org/wiki/Map_Features

c'est-à-dire aux grandes catégories regroupant plusieurs clés de tag (cf. section 3.2).

Autre page Tous les autres types de page du wiki qui ne tombent pas dans l'une des catégories sus-citées, comme, par exemple, les pages de profil des contributeurs OSM.

L'outil utilise l'information qu'il découvre sur le wiki OSM pour générer un graphe (l'ontologie OSN) dont les sommets (les concepts OSN) sont les pages Web de type *clé*, *tag* ou *cluster* et les arêtes (les relations OSN) sont les hyperliens entrants et sortants de ces mêmes pages Web. Bien que les pages de *proposition de tag*, de *cluster* ou les *autres pages* ne concernent pas directement une clé ou un tag OSM en particulier, leur contenu est indexé par le robot. En effet, ces ressources abordent de près ou de loin la folksonomie OSM et pointent vers des pages de type *tag*, *clé*, ce qui les rend exploitables du point de vue de la cooccurrence des références aux clés et tags.

Une fois la génération du graphe à partir de l'indexation des pages du Wiki OSM effectuée, une étape d'interconnexion des concepts OSN avec des bases de connaissances du Web Sémantique est opérée. En effet, des correspondances sont établies entre, d'un côté, les concepts extraits des pages du wiki OSM et, de l'autre, les pages de l'encyclopédie Wikipedia mais également les concepts de l'ontologie LGD. La mesure de similarité se base sur la proximité orthographique entre les chaînes de caractères que constituent les labels de chacun des deux concepts comparés. Les efforts des auteurs de [Auer et al., 2009] pour structurer les tags OSM sont ainsi conjointement mis à contribution avec le réseau OSN tissé sur des mesures de similarité entre ces mêmes tags.

3.7 Conclusion

Au cours de ce chapitre nous avons passé en revue l'ensemble des approches visant à structurer la folksonomie de la base de données OpenStreetMap. Nous avons vu comment le wiki OSM fait office de guide de bonnes pratiques relativement à l'usage des tags, mais également comment il assiste les contributeurs dans la délibération pour ou contre l'acceptation de nouveaux tags.

Dans un second temps, la méthode de conversion des tags OSM vers l'ontologie structurée *Linked GeoData* a été détaillée. Cette dernière distingue trois catégories de tags (attributs de classification, attributs de description et attributs de données), chacune faisant l'objet d'axiomes différents dans l'ontologie. Les travaux mis en

lumière dans [Auer et al., 2009] donnent ainsi lieu à une représentation des tags OSM dont l’expressivité formelle est supérieure à la simple folksonomie.

Ensuite, les principes de génération de l’ontologie OSMonto ont été exposés. Dans la lignée de l’ontologie LGD cette méthode de structuration des tags reprend le principe de hiérarchisation en deux niveaux, entre la clé et la valeur, des tags caractérisant la nature des objets géographiques de la base de données OSM. À cela, les auteurs de [Codescu et al., 2011] ajoutent la modélisation des relations de dépendances entre certains tags, donnant ainsi à voir la folksonomie OSM sous un autre jour.

Le fonctionnement de l’approche par jeu avec un but a , par la suite, été expliqué. Cette approche est à la fois ascendante en ce qu’elle met à contribution la communauté des utilisateurs OSM pour répondre à une série de questions pour la caractérisation des objets nouvellement cartographiés. Mais, l’approche proposée par les auteurs de [Baglatzi et al., 2012] est également descendante puisqu’elle cherche à aligner les tags OSM avec l’ontologie de haut niveau Dolce+DnS Ultra Lite afin de les rendre compatibles avec les standards du Web Sémantique.

Enfin, la méthode de construction du réseau sémantique OSN élaborée par les auteurs de [Ballatore et al., 2013] est présentée en fin de chapitre. Dans ce modèle, ce sont les cooccurrences des liens vers les pages de description des tags du wiki OSM qui sont collectées par un robot d’indexation. Le graphe qui en résulte rapproche les tags entre eux en leur associant un coefficient de similarité. Au-delà de l’interconnexion entre les concepts de l’ontologie OSN, des correspondances sont également établies entre les entités du graphe OSN et les classes de l’ontologie LGD, favorisant l’interopérabilité des efforts de structuration de la folksonomie OSM.

Dans le chapitre suivant, ces cinq modèles sont analysés à l’aune de quatre critères qui nous apparaissent indispensables pour la construction d’une ontologie des tags d’OpenStreetMap (OSM) : 1) la couverture de l’ensemble des tags, 2) l’interconnexion de l’ontologie avec d’autres bases de connaissances, 3) l’expressivité formelle de l’ontologie et 4) le respect de l’aspect participatif porté par le projet OpenStreetMap (OSM).

Chapitre 4

Limites des méthodes de structuration de la folksonomie OpenStreetMap

4.1 Introduction

Dans le chapitre 3, les cinq modèles de structuration de la folksonomie OpenStreetMap ont été passés en revue. Les méta-modèles et méthodes qui les sous-tendent ont été exposés. Selon les objectifs de chacun de ces efforts, un certain point de vue concernant l'interprétation des tags est adopté. D'abord, le site Web sous la forme d'un wiki est un premier pas vers une classification hiérarchique des tags. Les ontologies *Linked GeoData* [Auer et al., 2009] et OSMonto Codescu et al. [2011] insèrent des relations de subsomption entre les concepts respectivement extraits de la clé et de la valeur de tags décrivant la nature de l'objet géographique auquel ils sont associés. L'ontologie OSMonto établit une relation représentant la dépendance entre certains tags. Afin de faciliter la réutilisation de ces vocabulaires, les ontologies LGD et le réseau sémantique OSN [Ballatore et al., 2013] établissent respectivement des correspondances avec les bases de connaissances DBpedia¹ et WordNet². Toujours dans l'optique d'encourager l'interopérabilité de ces ontologies, le GWAP [Baglatzi et al., 2012] se propose d'aligner les tags OSM avec une l'ontologie de haut niveau DUL par l'intermédiaire d'un questionnaire adressé aux contributeurs.

Dans ce chapitre, la focale est mise sur quatre aspects qui nous semblent cru-

1. <http://dbpedia.org/>

2. <https://wordnet.princeton.edu/>

ciaux pour la construction d'une ontologie des tags OSM. Le premier d'entre eux, évoqué en section 4.2, concerne la couverture de la folksonomie, c'est-à-dire le degré d'inclusion des tags dans le modèle proposé. Le second critère, discuté en section 4.3, a trait à la connexion du modèle avec d'autres bases de connaissances. La section 4.4 traite de l'expressivité formelle de la structure dans laquelle sont représentés les tags. L'aspect participatif, cher au projet OSM, est examiné en section 4.5. Finalement, en section 4.6, les cinq modèles de structuration de la folksonomie OSM abordés dans le chapitre 3 sont analysés sur la base de ces quatre critères afin d'identifier les limites auxquelles nous proposons des réponses dans la partie II de cette thèse.

4.2 Une couverture lacunaire des tags

Selon le wiki OSM, il existe aujourd'hui 77 millions de tags³ dans la base de données OSM. D'après [Ballatore et al., 2013], en 2013, ce wiki regroupait 2047 tags. Les auteurs de [Auer et al., 2009] se basent sur cet ensemble de tags pour générer l'ontologie *Linked GeoData*. De leur côté, les auteurs de [Codescu et al., 2011] prennent non seulement en compte les tags issus de la page Map Features⁴ du wiki OSM, mais également les tags qui présentent au moins 100 occurrences dans la base de données. En définitive, seul l'approche par GWAP adopte une approche potentiellement capable de considérer tous les tags. Toutefois, la méthode semi-automatique GWAP ne lie un tag avec l'ontologie de haut niveau DUL que lorsqu'un contributeur l'associe avec un objet géographique. Par conséquent, les tags existants dont l'ajout est antérieur à la mise en place de cette approche sont laissés de côté.

Pourtant, en écartant une partie significative des contributions, OSMonto est privée d'une grande diversité de contributions minoritaires potentiellement intéressantes pour saisir la représentation du territoire d'un point de vue citoyen. En effet, même si, dans certains cas, il s'agit bien d'erreurs d'orthographe et autres fautes de frappe, le fait d'écarter des tags sur la base de leur nombre d'occurrences risque d'entraver l'émergence de tags potentiellement pertinents, et, par là-même, de porter atteinte à un atout majeur de l'activité de tagging : son aspect évolutif. Pour vérifier cette hypothèse, et dans la continuité des travaux de M. Dahl [Dahl, 2015a,b], nous envisageons d'étudier les profils d'émergence des tags OSM en s'appuyant sur l'historique de leurs modifications⁵.

3. <http://wiki.openstreetmap.org/wiki/Taginfo/FAQ>

4. http://wiki.openstreetmap.org/wiki/Map_Features

5. L'API Overpass permet, via l'URL http://overpass-api.de/augmented_diffs/, d'obtenir la liste des modifications dans la base de données OSM survenues entre deux dates.

4.3 Une interconnexion incomplète

Les modèles de données élaborés à partir de la folksonomie OSM, évoqués dans le chapitre 3, expriment des points de vue différents quant à l'interprétation des tags. Par exemple, d'un côté, l'ontologie LGD établit des relations de subsomption (`rdfs:subClassOf`) entre les clés et les valeurs des tags de type *attribut de classification* : le tag `shop=bakery` est interprété de telle sorte que le concept `Bakery` est une forme particulière (sous-classe, `rdfs:subClassOf`) du concept `Shop`. De l'autre côté, le réseau sémantique OSN exprime des relations de similarité (`owl:equivalentClass`) entre ces mêmes tags. De plus, les modèles pour la structuration de la folksonomie OSM utilisent des concepts faisant référence à des choses distinctes. Par exemple, l'ontologie OSMonto ne représente que les tags dont la valeur est "une constante OSM plutôt qu'une chaîne de caractères ou un nombre" tandis que OSN considère également des concepts représentant des pages *cluster* du wiki OSM, le contenu de ces dernières n'évoquant pas directement un tag en particulier, mais bien une catégorie thématique recouvrant plusieurs tags (cf. section 3.2).

Cette hétérogénéité des concepts et relations est le reflet de postulats différents, mais potentiellement complémentaires, dans la conception des modèles de structuration de la folksonomie OSM. Les auteurs de [Ballatore et al., 2013] abondent dans ce sens : l'ontologie OSN dresse des correspondances entre ses propres concepts, ceux du wiki OSM en les indexant et ceux de l'ontologie LGD. Cependant, l'expression des dépendances entre tags élaborée par les auteurs de [Codescu et al., 2011] ne peut pas être exploitée conjointement avec les autres modèles de structuration de tags OSM car l'ontologie OSMonto n'a été interconnectée avec aucune autre ontologie du Web Sémantique. Dans l'optique de profiter du maximum d'informations disponibles sur les tags, il est nécessaire de construire des ponts entre les différentes ontologies de tags OSM dont les représentations sont complémentaires. Par exemple, lors de la consultation d'un objet cartographié dans la base de données OSM, un utilisateur voudra aussi bien connaître le descriptif des tags associés à l'élément sélectionné (via les informations disponibles sur les pages correspondantes du wiki OSM), l'organisation hiérarchique si elle existe entre ces tags (via l'ontologie LGD), les relations de dépendances entre tags si elles existent (via l'ontologie OSMonto), ainsi que le degré de similarité qu'ils partagent (via le réseau sémantique OSN).

Si cet aspect complémentaire des informations exprimé dans différentes représentations des tags OSM doit être pris en compte, il existe également des divergences entre certains modèles. Pour illustrer ce cas d'incompatibilité, prenons l'exemple de OSMonto qui, lors de la création de la hiérarchie clé-valeur, traite les tags dont la valeur est booléenne de la même manière que les tags qui ont pour valeur des

"constantes OSM" [Codescu et al., 2011]. Ainsi, le `wheelchair=yes` produira une taxonomie à deux niveaux, le concept `Wheelchair` subsumant le concept `Yes`. Au contraire, l'ontologie LGD prend le parti de distinguer les tags de type *attribut de classification* correspondant à la catégorie "Caractéristique Principale" du wiki OSM (cf. section 3.2) et les tags de type *attribut de données* dont les valeurs sont des types de données primitifs (chaînes de caractères, nombre, date, booléens, tout ce qui peut être utilisé comme rang d'une propriété `owl:DatatypeProperty`, c'est-à-dire toutes valeurs représentées dans le *XML Schema Definition (XSD)*⁶). Contrairement au paradigme OSMonto, l'ontologie LGD ne crée pas de concept pour le tag `wheelchair=yes`, mais une propriété objet (`owl:ObjectProperty`). Rendre interopérables des ontologies fondées sur des hypothèses divergentes en conservant la diversité des points de vue pourrait s'avérer utile pour améliorer l'acquisition de la données OSM en désambiguïsant les tags.

Dans l'ontologie LGD, les instances (c'est-à-dire la ABox, en l'occurrence, les objets cartographiés de la base de données OSM) sont liées à l'encyclopédie Wikipédia et, par conséquent, à la base de connaissances DBpedia puisque les URLs Wikipédia sont les identifiants des entités de DBpedia. Par exemple, l'URL du domaine Wikipédia http://en.wikipedia.org/wiki/Frauenkirche_Dresden est associée au noeud représentant l'église Notre-Dame de Dresde. Cependant, les classes et les relations (c'est-à-dire la TBox, en l'occurrence les tags OSM) issues de cette même ontologie ne sont directement connectées à aucune autre base de connaissances. De son côté, le réseau sémantique OSN bâtit des liens entre les concepts extraits des tags et ceux de la base de connaissances WordNet⁷.

En revanche, comme les auteurs de [Ballatore et al., 2013] l'ont souligné, des connexions doivent également être établies entre des bases de données géospatiales telles que GeoWordNet⁸ ou Geonames⁹ qui font aujourd'hui autorité et OSN. Par exemple, au moment de la modification de la base de données OSM, lorsqu'un utilisateur souhaite ajouter un certain tag à un élément spatialisé, s'il a la possibilité de profiter de la définition précise, en provenance du réseau lexical WordNet, qui est liée à ce tag, il sera soit conforté dans sa décision si la définition correspond avec ce qu'il avait en tête concernant l'interprétation du tag, ou bien, au contraire, dissuadé de le faire si la caractérisation du tag est liée à des notions qui ne sont pas celles qu'il voulait exprimer au moyen de ce tag.

6. <https://www.w3.org/TR/xmlschema-2/>

7. <https://wordnet.princeton.edu/>

8. <http://datahub.io/dataset/geowordnet>

9. <http://www.geonames.org/>

4.4 Une structure à l’expressivité limitée

Parmi les modèles vus au chapitre 3, seules des relations de subsumption, de similarité et d’instanciation sont exprimées directement dans les ontologies. Nos travaux se concentrent sur les tags, c’est pourquoi nous mettons de côté la relation d’instanciation entre tags et objets géographiques exprimée dans l’ontologie LGD. L’ontologie issue de l’approche GWAP hérite de la hiérarchie de l’ontologie haut-niveau DUL, mais ne génère pas de relations de subsumption entre les tags. Le réseau sémantique OSN hérite des relations de méronymie (le rapport du tout à ses parties), antonymie (le rapport de symétrie sémantique) et troponymie (le degré d’intensité) de WordNet, mais ne cherche pas à en créer de nouvelles entre les tags. De plus, l’approche adoptée par OSN extrait des relations de similarité entre les tags basée, entre autres, sur la structure de l’ontologie LGD. Cette dernière, à l’instar de l’ontologie OSMonto, intègre des relations de subsumption entre les concepts extraits des tags OSM.

Ce rapport hiérarchique n’intervenant qu’entre la clé et la valeur d’un tag, la taxonomie générée, présentée en figure 4.1 est très simpliste : elle ne compte que deux niveaux. Cette superficialité est un obstacle pour l’automatisation de certaines tâches. Par exemple, au moment de l’acquisition, lorsqu’un contributeur ajoute un tag à la base de données OSM, un système de suggestion de tags similaires pourrait être mis à profit dans le but d’améliorer la précision du tag choisi. Il existe des outils pour mesurer le degré de similarité entre des concepts d’une ontologie (comme l’ontologie LGD, par exemple) : les *matchers* d’ontologies.

Afin de trouver des similitudes entre les classes, les *matchers* utilisent différentes techniques (basées sur le nom des concepts, sur la structure de l’ontologie, sur des ressources externes, *etc.*) et stratégies (chaînage de plusieurs *matchers*, aggrégation des similarités, méthodes d’apprentissage, implication de l’utilisateur, *etc.*). J. Euzenat et S. Pavel fournissent un état de l’art complet à ce sujet [Euzenat and Shvaiko, 2013]. Chaque correspondance est établie entre deux concepts par un degré de similarité. L’ensemble de ces correspondances est appelé un *alignement*. Certaines techniques de *matching* consistent en l’exploitation de l’expressivité formelle de la *structure* de l’ontologie, c’est-à-dire les relations qui permettent d’agencer les concepts entre eux. La relation de subsumption en est un exemple.

Ainsi, dans une ontologie structurée de manière hiérarchique, deux concepts frères qui partagent le même parent via un lien de subsumption se verront attribuer un degré de similarité élevé par un *matcher* qui adopte une stratégie d’analyse structurelle. Cette stratégie est d’autant plus efficace que la hiérarchie considérée est

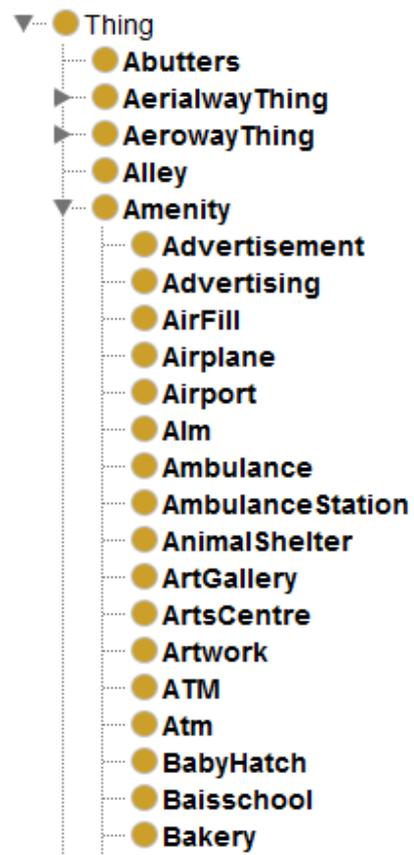


FIGURE 4.1 – Aperçu de la hiérarchie de tags issue de l'ontologie LGD.

profonde. Les ontologies LGD et OSMonto ne possédant que deux niveaux, cette approche demeure donc limitée. Par conséquent, lorsqu'un contributeur OSM souhaite ajouter le tag `amenity=school`, le système de suggestion utilisant les correspondances produites par une stratégie d'alignement basée sur la hiérarchie clé-valeur de LGD ou OSMonto propose des tags frères, c'est-à-dire ayant la clé `amenity`. Ainsi, les tags `amenity=bar`, `amenity=parking` ou encore `amenity=dentist` partagent le même degré de similarité avec le `amenity=school`, malgré une distance sémantique dont l'importance apparaît comme évidente pour un jugement humain. Cette superficialité de la taxonomie est donc un frein majeur au contrôle de la qualité des données au moment de l'acquisition.

L'ontologie LGD propose la typologie de tags OSM la plus fine des modèles vus en chapitre 3. En effet, les tags y sont organisés en trois catégories : les attributs de *classification*, les attributs de *description* et les attributs de *données*. Cependant, la raison pour laquelle les attributs de classification font l'objet d'une *classe* (`owl:class`) dans l'ontologie alors que les attributs de description et attributs de données sont convertis en *propriété* (respectivement `owl:ObjectProperty` et `owl:DataProperty`) est discutable. En effet, de prime abord, des tags tels que `amenity=restaurant` semblent décrire la nature de l'objet auquel ils sont associés, tandis que des tags tels que `wheelchair=yes` ne représente qu'une propriété de l'objet.

Toutefois, si l'on considère l'utilisation conjointe des tags `amenity=cafe` et `internet_access=yes` dans le cas de la description d'un cybercafé, on voit mal lequel des deux est le plus pertinent pour caractériser la nature intrinsèque de l'objet. Pourtant, seules les classes `Amenity` et `Cafe` sont générées dans LGD tandis qu'aucune classe `InternetAccess` n'est créée. L'usage ambigu de certains autres tags témoigne également de cette lacune : le tag `building=yes`, très largement utilisé par la communauté OSM¹⁰, est classifié comme un attribut de données par l'approche des auteurs de [Auer et al., 2009] : aucun concept `Building` n'est donc généré à partir de ce tag. Pourtant, tous les autres tags partageant la clé `building` (`building=church`, `building=stadium`, `building=kiosk`, *etc.*), sont considérés comme des attributs de description : cette fois le concept `Building` sera ajouté à l'ontologie. L'exploitation d'un maximum de tags dans le but de construire une ontologie plus étoffée permettrait, par exemple, de faciliter l'exploration des données.

10. Il est associé à plus de 4% de l'ensemble des objets en base de données OSM selon le service Web TagInfo, <http://taginfo.openstreetmap.org/tags/building=yes>

4.5 Une implication faible des contributeurs OSM

Enfin, aussi bien les auteurs de [Codescu et al., 2011], à l’origine de l’ontologie OSMonto, que ceux de [Ballatore et al., 2013], à l’origine du réseau sémantique OSN, s’accordent à dire que les méthodes et outils d’alignement automatique des concepts ne sont pas suffisants. Pour pallier ces approximations, une expertise humaine est nécessaire dans le but d’interconnecter efficacement les concepts extraits des tags OSM avec d’autres bases de connaissances du Web Sémantique. Une telle approche semi-automatique est cohérente avec le projet de cartographie par et pour les contributeurs OSM.

Dans l’approche préconisée par les auteurs de [Baglatzi et al., 2012], les utilisateurs ont un rôle central. En effet, ce sont leurs réponses aux questions liées à l’ontologie de haut niveau DUL qui permettent de connecter les tags à ceux de l’ontologie DUL. Cependant, l’étape de traduction des classes qui composent l’ontologie DUL est réalisée par des experts et n’est pas soumise à un processus participatif. De plus, la latitude laissée au contributeur dans l’interprétation de chaque question se répercute nécessairement dans l’alignement. Cette couche d’expertise confère à cette méthode un aspect descendant qui va, dans une certaine mesure, à l’encontre de la philosophie participative du projet OpenStreetMap.

4.6 Vers un nouveau modèle pour la sémantisation des tags OSM

Le tableau 4.1 livre une synthèse des quatre modèles de données pour les tags OSM étudiés dans le chapitre 3 évalués sur la base des quatre critères suivants : 1) exhaustivité de la couverture des tags, 2) interconnexion avec les bases de connaissances existantes, 3) implication de l’utilisateur, 4) expressivité formelle. Plusieurs carences propres à la sémantique relative à l’Information Géographique Volontaire (*Volunteered Geographic Information*, VGI) sont traitées par ces modèles. En effet, le GWAP est potentiellement capable d’intégrer tous les tags d’OSM. Le wiki OSM, quant à lui, implique de façon significative les utilisateurs dans le choix des tags et gère les différentes versions des pages du site web. Enfin, le réseau sémantique OSN est un premier pas vers une interconnexion substantielle entre les concepts issus de la folksonomie OSM et ceux qui proviennent d’autres bases de connaissances, tout en offrant une taxonomie de tags relativement expressive grâce à des mesures de similarité entre ces concepts. Cependant, aucune de ces contributions ne répond à l’ensemble des quatre critères. C’est pourquoi nous proposons un modèle pour assis-

	COUVERTURE	INTERCONNEXION	EXPRESSIVITÉ	IMPLICATION UTILISATEUR
OSM Wiki	Faible • 2047/70M tags	Faible • Hyperliens Wikipedia	Aucune	Forte • Cycle BRD
LGD	Faible • 2047/70M tags	Moyenne • DBpedia	Faible • Subsumption	Faible • Expertise descendante
OSMonto	Moyenne • (2047 + tags _{occurrences>100})/70M tags	Aucune	Faible • Subsumption	Faible • Expertise descendante
GWAP	Forte • Potentiellement tous	Faible • DUL	Moyenne • Subsumption (DUL) • Similarité (DUL)	Moyenne • Questionnaire par expert • Caractérisation par utilisateur
OSN	Faible • 2047/70M tags	Forte • LGD • DBpedia • WordNet	Moyenne • Subsumption (LGD) • Méronymie (WordNet) • Antonymie (WordNet) • Troponymie (WordNet) • Similarité	Faible • Ontologies descendantes

TABLE 4.1 – Comparaison des modèles de données de données pour les tags OSM.

ter la construction collaborative d'une ontologie des tags OSM faite par et pour ses utilisateurs, qui tire le meilleur parti des travaux précédents et satisfasse l'intégralité des besoins évoqués plus haut.

Notre approche vise à assister la construction d'un modèle conceptuel pour OSM (i.e. une ontologie) capable de couvrir la sémantique contenue dans la folksonomie d'OSM et qui réponde à plusieurs problèmes. En premier lieu, afin de tenir compte de tous les éléments cartographiques (c'est-à-dire les représentations des objets physiques présents sur le territoire tels que les espaces verts, les équipements publics, les commerces, le réseau routier, *etc.*) qui sont enregistrés dans la base de données, l'ontologie qui les modélise doit intégrer tous les tags existants qui les décrivent. La couverture des tags est donc un aspect de la plus haute importance pour prendre en considération la sémantique d'OSM de façon exhaustive. D'autre part, l'un des intérêts principaux de la création d'une ontologie est la réutilisation d'un vocabulaire commun dans l'optique de favoriser l'interopérabilité des systèmes informatiques. En cela, l'interconnexion de l'ontologie OSM avec les autres bases de connaissances qui existent dans l'écosystème des Données Ouvertes et Liées (*Linked OpenData*, LOD)¹¹ est essentielle.

D'autre part, la base de connaissances OSM dont nous nous proposons d'appuyer la construction ne pourra être efficacement exploitée que si elle est solidement structurée. Dès lors, le choix d'une sémantique pertinente doit garantir l'expressivité des types de concepts et relations décrits dans le méta-modèle qui la sous-tend. Finalement, dans la lignée des préceptes posés par le projet OSM, nous faisons le pari que la confiance en l'intelligence collective [O'Reilly, 2005] est cruciale pour encourager la responsabilité et l'autonomie des citoyens dans le processus de production de l'Information Géographique. Une attention toute particulière sera donc accordée à l'implication de l'utilisateur dans le modèle que nous proposons.

4.7 Conclusion

Au travers de ce chapitre, nous avons soulevé quatre aspects qui nous paraissent essentiels en vue de structurer la folksonomie OSM. Le premier d'entre eux, abordé en section 4.2, insiste sur l'intérêt d'adopter une approche globale dans la considération de l'ensemble des tags OSM. L'approche GWAP adoptée par les auteurs de [Baglatzi et al., 2012] est certainement celle qui parvient le mieux à respecter cet aspect.

11. <http://lod-cloud.net/>

En section 4.3, c'est l'interconnexion des concepts extraits des tags OSM avec d'autres bases de connaissances qui est discutée. La construction de liens entre les ontologies de tags, mais également avec d'autres ontologies du Web sémantique, est cruciale dans l'optique de les rendre interopérables. Cet aspect présente le double intérêt de profiter du maximum d'informations disponibles sur un concept donné, mais également de confronter les points de vue qui sont à l'origine d'ontologies différentes. En la matière, le réseau sémantique OSN propose des connexions vers l'ontologie LGD mais également vers le réseau lexical WordNet.

Ensuite, dans la section 4.4, les avantages d'une ontologie dont l'expressivité formelle est importante sont expliqués. Une ontologie de tags dont les relations sont trop peu expressives peut être un obstacle majeur s'agissant des possibilités de raisonnement sur son graphe. C'est tout particulièrement vrai dans le cas de l'utilisation des briques technologiques que sont les matchers d'ontologie.

Par ailleurs, la fidélité à la philosophie du projet OSM, quant à l'implication des contributeurs, est soulignée en section 4.5. Le wiki OSM ainsi que la méthode semi-automatique GWAP sont les efforts qui sont les plus cohérents avec cet aspect participatif.

Enfin, en section 4.6, chacun des cinq modèles vus dans le chapitre 3 a été l'objet d'une analyse à travers le prisme de ces quatre critères. Ceci nous a permis d'énoncer les objectifs de notre contribution qui vise l'élaboration d'un modèle pour la structuration des tags associés aux objets cartographiés dans le cadre du projet OSM.

Deuxième partie

Contributions

Chapitre 5

Le méta-modèle de OF4OSM

5.1 Introduction

Dans la partie I, nous avons proposé un panorama des efforts réalisés pour remédier aux inconvénients inhérents à la folksonomie du projet OpenStreetMap (OSM). Les principes d’OSM ont été présentés dans le chapitre 2 et les modèles qui visent à en structurer les tags (OSM wiki, *Linked GeoData* (LGD), OSMonto, GWAP et *OSM Semantic Network* (OSN)) ont été passés en revue dans le chapitre 3.

Dans le chapitre 4, nous avons montré les limites des modèles pour la structuration de la folksonomie d’OSM. Ces derniers ont été analysés sur la base de 4 critères : 1) la couverture de l’ensemble des tags (section 4.2), 2) l’interconnexion du modèle avec d’autres bases de connaissances (section 4.3), 3) l’expressivité formelle du modèle, c’est-à-dire la richesse des relations qu’il permet d’exprimer (section 4.4), 4) l’implication de l’utilisateur dans le processus d’élaboration de la base de tags (section 4.5). Chaque modèle présente des atouts au vu de certains critères, mais aucun ne parvient à satisfaire tous les critères.

Dans ce chapitre, nous présentons le méta-modèle qui structure l’Ontologie extraite de la Folksonomie pour OSM (*Ontology from Folksonomy for OSM*, OF4OSM). La figure 5.1 représente les deux étapes du processus de génération de l’ontologie OF4OSM. La première étape consiste en la caractérisation des tags OSM au moyen d’un modèle de représentation décrit en section 5.2. La seconde étape correspond à la structuration des tags précédemment caractérisés en une ontologie. Cette structuration s’appuie sur un modèle conceptuel présenté dans la section 5.3. Finalement, la section 5.4 résume les apports du méta-modèle de OF4OSM abordés au cours de

ce chapitre.

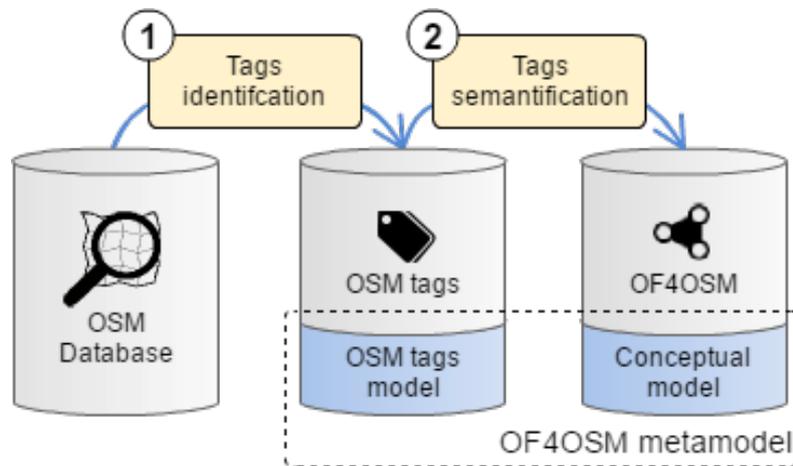


FIGURE 5.1 – Processus de génération de l’ontologie OF4OSM appuyé sur un meta-modèle en deux parties : un modèle de représentation des tags OSM et un modèle conceptuel.

5.2 Représentation des tags OSM dans OF4OSM

5.2.1 Une représentation conforme aux standards

La plupart des plates-formes de Contenu Généré par les Utilisateurs (*User Generated Content*, UGC) utilise des tags pour classifier leur contenu. Notre approche s’insère dans ce courant et vise à améliorer l’expressivité formelle des folksonomies en structurant les tags en une ontologie. Dans [Lohmann et al., 2011], les auteurs font un état de l’art des ontologies de tags et proposent l’ontologie unifiée de l’activité de *tagging* (*Modular Unified Tagging Ontology*, MUTO)¹ qui vise à unifier les concepts fondamentaux sur lesquels repose l’activité de tagging et que l’on retrouve dans différentes ontologies qui font autorité telles que *Tag Ontology*², *Meaning Of A Tag* (MOAT)³, *Common Tag*⁴ ou *NiceTag*⁵. L’ontologie MUTO fait référence à des vocabulaires tels que *Semantically-Interlinked Online Communities* (SIOC)⁶ qui permet de lier un utilisateur qui utilise des tags (un *tagger*) à une communauté en ligne. MUTO réutilise également les schémas de métadonnées DCTERMS

1. <http://purl.org/muto/core>
2. <http://www.holygoat.co.uk/owl/redwood/0.1/tags/>
3. <http://moat-project.org/ns>
4. <http://commontag.org/ns>
5. <http://ns.inria.fr/nicetag/2010/09/09/voc.rdf>
6. <http://rdfs.org/sioc/ns>

maintenus par l'initiative Dublin Core pour les méta-données⁷ dédiées à la description formelle et thématique des jeux de données. Le langage de représentation des connaissances SKOS⁸, recommandé par le *World Wide Web Consortium* (W3C)⁹, est lui-aussi sollicité dans MUTO. Il permet d'exprimer des relations sémantiques entre concepts ainsi que des correspondances avec d'autres schémas de représentation des connaissances du LOD. La représentation des tags dans le respect de ces différents standards rend interopérable l'ontologie OF4OSM avec les bases de connaissances du Web Sémantique.

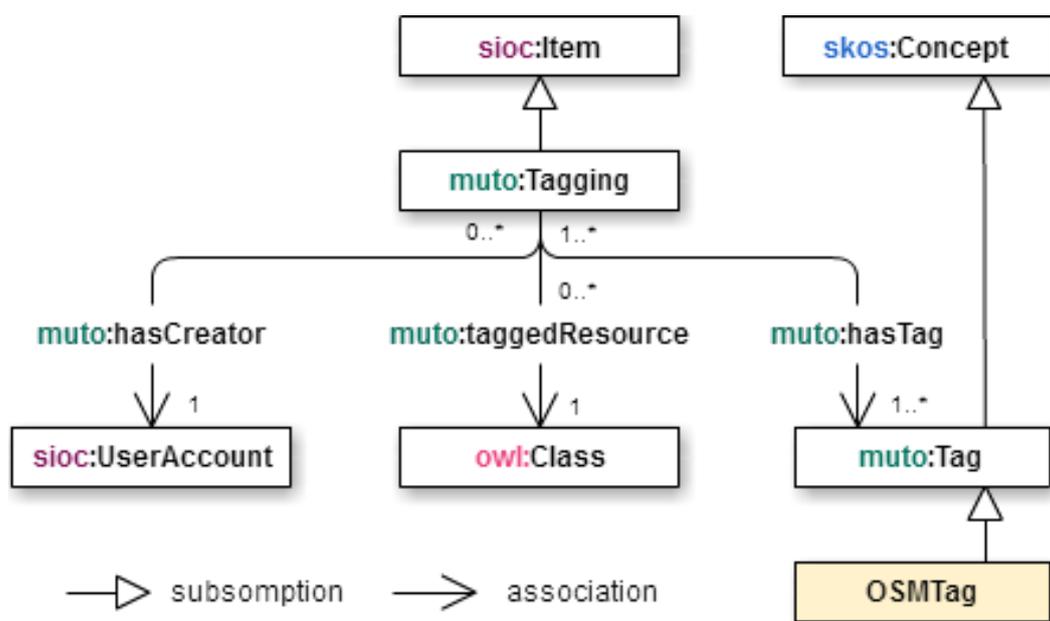


FIGURE 5.2 – Positionnement du concept `of4osm:OSMTag` au sein des vocabulaires standards pour la représentation de l'activité de *tagging*.

La figure 5.2 illustre le positionnement du concept `of4osm:OSMTag`, qui représente un tag OSM dans notre approche, par rapport à l'ontologie de l'activité de *tagging* MUTO et aux vocabulaires auxquels elle fait référence. L'ontologie MUTO est ici présentée de façon simplifiée pour plus de lisibilité. Un schéma exhaustif de ses composantes et relations est disponible en annexe A. L'activité de *tagging* `muto:Tagging` est une spécialisation du concept `sioc:Item` qui permet de décrire un contenu au sens large qui fait l'objet d'interaction entre les membres d'une communauté. Le concept `muto:Tagging` est lié à 1) un *tagger* dont le compte est modélisé par le concept `sioc:UserAccount` via la propriété `muto:hasCreator`; 2) une ressource (`owl:Class`) via la propriété `muto:taggedResource`; 3) un ou plu-

7. <http://purl.org/dc/terms/>

8. <https://www.w3.org/2009/08/skos-reference/skos>

9. <http://www.w3.org/>

sieurs tags (`muto:Tag`) par la propriété `muto:hasTag`. Afin de tirer le meilleur parti de ces standards, `of4osm:OSMTag` est déclaré sous-concept de `muto:Tag`. De plus, puisque `muto:Tag` est aussi un `skos:Concept`, `of4osm:OSMTag` est, par transitivité, lui-aussi décrit dans ce langage de représentation des connaissances largement utilisé en sciences de l'information. Le langage SKOS permet d'établir des correspondances entre des concepts équivalents exprimés dans d'autres schémas grâce aux relations `skos:closeMatch` et `skos:exactMatch` dans le but d'encourager l'interconnexion des ontologies du Web Sémantique.

Toujours dans l'optique d'optimiser l'interopérabilité avec les standards de la représentation des connaissances et du Web, mais également parce que nous avons besoin d'un formalisme qui supporte des relations de subsomption (cf. sections 5.2 et 5.3), le méta-modèle de OF4OSM est décrit dans le *Ontology Web Language*¹⁰. OWL est un langage de représentation des connaissances construit sur le modèle de données de *Resource Description Framework* (RDF)¹¹. Ce dernier est destiné à décrire de façon formelle les ressources du Web et leurs métadonnées notamment sous la forme d'un graphe afin d'en automatiser les traitements. C'est le standard de base du Web sémantique. Le formalisme RDF permet d'exprimer les données sous la forme d'associations ternaires, chacune impliquant un sujet, un prédicat et un objet : on parle alors de "triplets RDF". OWL s'appuie sur ce standard : le sujet du triplet RDF est exprimé par le concept `owl:Class`, le prédicat est représenté soit par le concept `owl:DataProperty` si l'objet est un type de données simple (entier, booléen, chaîne de caractères, *etc.*), soit par le concept `owl:ObjectProperty` si l'objet est une autre instance du concept `owl:Class`. De plus, OWL permet de décrire les ontologies du Web au moyen d'axiomes de logique de description (équivalence, restriction, union, intersection, complément, subsomption, *etc.*), dans le but d'inférer de nouvelles connaissances à partir d'une base de données. Le schéma du méta-modèle OWL est donné en annexe B.

Ainsi, chacun des concepts présents dans OF4OSM est une instance du concept `owl:Class` et chacune des relations entre ces concepts est soit une instance de la classe propriété d'objet `owl:ObjectProperty` si son image est un autre concept, soit une instance de la classe propriété de type de données `owl:DataProperty` si son image est de type simple (`xsd:String`, `xsd:Integer`, `xsd:Decimal`, `xsd:Boolean`, `xsd>Date`, `xsd:Time`, *etc.*^{12 13}). La compatibilité entre les langages OWL et SKOS est assurée par des règles de conversion formalisées par le W3C¹⁴.

10. <https://www.w3.org/TR/owl-features/>

11. <https://www.w3.org/RDF/>

12. <https://www.w3.org/TR/swbp-xsch-datatypes/>

13. http://www.w3schools.com/xml/schema_simple.asp

14. <https://www.w3.org/2006/07/SWD/SKOS/skos-and-owl/master.html>

5.2.2 Représentation générale d'un tag OSM

Malgré tous ses avantages, le modèle de représentation d'un tag établi par l'ontologie MUTO ne correspond pas exactement à la structure d'un tag OSM représenté par une paire clé-valeur (cf. chapitre 2). Par conséquent, dans le méta-modèle de OF4OSM, un tag, représenté par le concept `of4osm:OSMTag`, possède deux sous-parties, une clé, représentée par le concept `of4osm:OSMTagKey`, et une valeur, représentée par le concept `of4osm:OSMTagValue`. Ces dernières sont liées au concept `of4osm:OSMTag` par la propriété méréologique `dc:hasPart`. Afin de rendre notre modélisation facilement réutilisable, les concepts du méta-modèle de OF4OSM font l'objet d'une API dans laquelle les noms des interfaces sont préfixés par un `I`. Ainsi, les classes `of4osm:OSMTag`, `of4osm:OSMTagKey` et `of4osm:OSMTagValue` implémentent respectivement les interfaces `of4osm:IOSMTag`, `of4osm:IOSMTagKey` et `of4osm:IOSMTagValue`.

Outre ces deux composantes majeures, le concept `of4osm:IOSMTag` possède un attribut `wikiURL` indiquant, si elle existe, l'URL de la page du Wiki OSM correspondant au tag, conservant ainsi une information précieuse quant à l'autorité du tag (le wiki OSM est un guide des bonnes pratiques relatives à l'utilisation des tags dans OSM, cf. section 3.2), et à partir de laquelle il est aisé de retrouver la sémantique associée au tag. Le concept `of4osm:IOSMTagKey` possède également un attribut `wikiURL` pour des raisons identiques puisqu'il existe des pages sur le Wiki OSM qui sont dédiées aux clés.

De plus, nous avons classé les deux sous-éléments du tag, la clé et la valeur, en plusieurs types. En effet, après analyse de la page du Wiki OSM Map Features¹⁵ qui décrit l'utilisation des tags OSM les plus consensuels, nous avons distingué 4 types de clé (`of4osm:IOSMTagKey`) possibles et 6 types de valeur (`of4osm:IOSMTagValue`) possibles. Cette caractérisation des tags OSM est intégrée dans le méta-modèle OF4OSM sous la forme d'une extension à l'ontologie MUTO présentée sur la figure 5.3.

5.2.3 Représentation des différents types de clé

Les clés des tags OSM (`of4osm:IOSMTagKey`) se déclinent sous différentes formes que nous avons regroupées en 4 types (partie gauche de la figure 5.3). Chacun de ces types correspond à une manière d'interpréter la valeur qui lui est associée. Les 4 types de clés sont les suivants :

15. http://wiki.openstreetmap.org/wiki/Map_Features

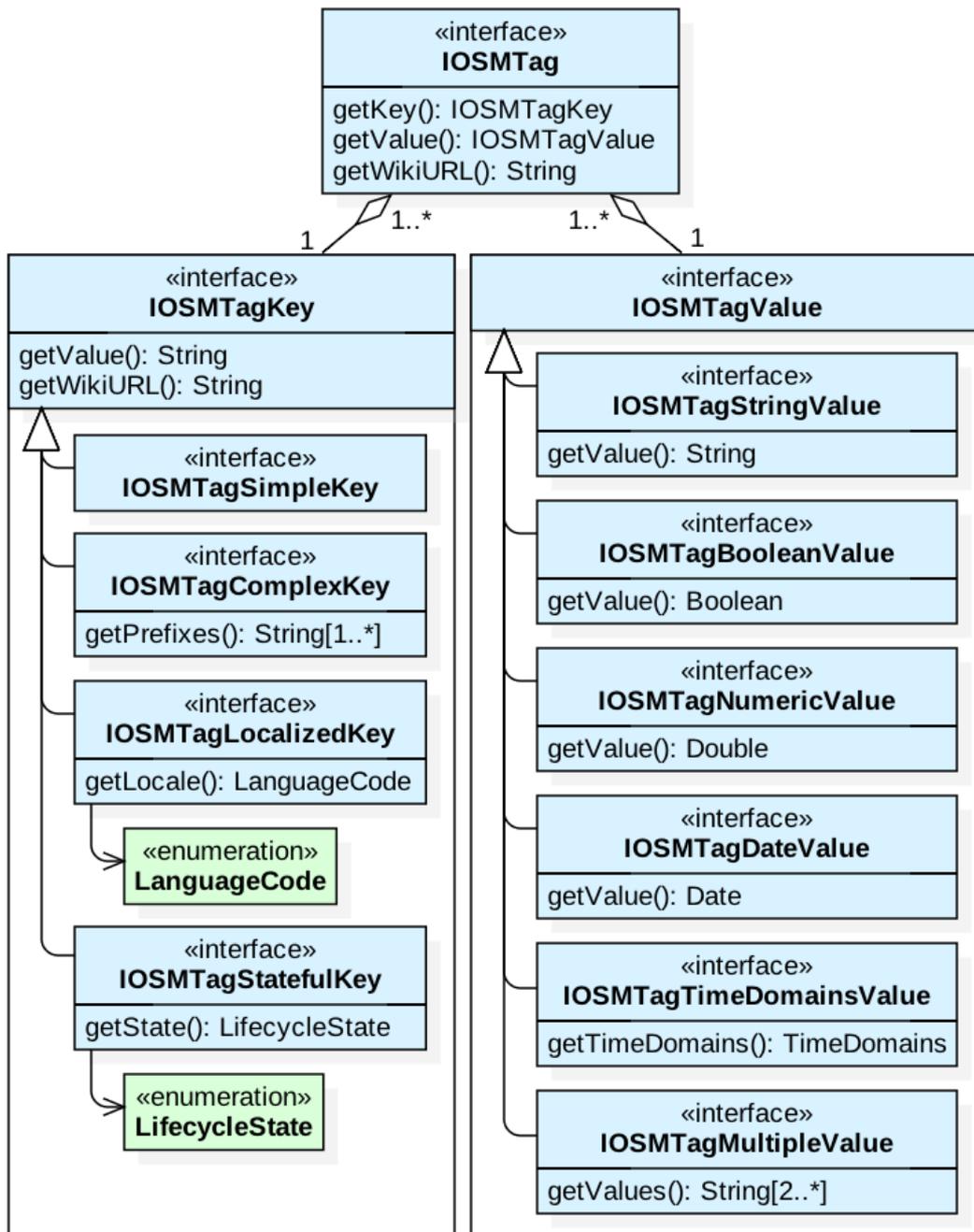


FIGURE 5.3 – Modèle de représentation d'un tag OSM.

Clé simple (`of4osm:IOSMTagSimpleKey`) Les clés simples sont des clés qui ne sont composées que d'un seul lemme, c'est-à-dire d'une seule unité lexicale mais pouvant être constituées d'un ou plusieurs morphèmes séparés par le signe typographique tiret bas (représenté par le symbole "_"). Par exemple, `amenity`, `shop`, `building` (un seul morphème) ou encore `man_made` et `social_facility` (deux morphèmes). Ces clés peuvent traduire une propriété (`height`, `name`) ou bien une catégorie (`amenity`, `shop`);

Clé complexe (`of4osm:IOSMTagComplexKey`) Les clés complexes sont des clés qui comprennent plusieurs lemmes séparés par un deux-points (représenté par le symbole ":"). Par exemple, `addr:housenumber`, `building:levels` ou encore `climbing:grade:british:technical`. Pour ce type de clé, l'ordre des lemmes est significatif : chaque lemme est spécialisé par le lemme suivant. Pour illustrer ce point, prenons l'exemple de la clé `climbing:grade:british:technical` : les valeurs associées à cette clé complexe concernent l'activité d'escalade (`climbing`), plus spécifiquement les cotations des voies (`grade`) conformément aux standards britanniques (`british`) et selon les nomenclatures techniques (`technical`). Ce type de clé représente une propriété. Cette propriété peut être rattachée à un tag qui a valeur de catégorie, le lien entre les deux se fait grâce à la correspondance entre le premier lemme de la clé complexe d'un côté et la valeur du tag de l'autre. Dans le cas de la pratique de l'escalade, il s'agit du tag `sport=climbing` qui décrit les aménagements dédiés à cette activité.

Clé internationalisée (`of4osm:IOSMTagLocalizedKey`) Les clés internationalisées sont constituées de deux lemmes séparés par un deux-points (:). Le premier est une propriété qui attend une valeur sous la forme d'une chaîne de caractères tandis que le second précise le langage (via le code langage ISO 639.2¹⁶) dans lequel est exprimée la valeur associée. À titre d'exemple, le tag `name:fr=Londres` indique que, pour la clé `name`, la valeur donnée ici (*Londres*) est la version française (code `fr`) du signifiant faisant référence à l'entité géographique dénommée *London* dans la langue officielle du pays dont elle est la capitale.

Clé avec état (`of4osm:IOSMTagStatfulKey`) Les clés avec état sont elles-aussi composées de deux lemmes. Le premier se rapportant à une catégorie permettant de caractériser un objet géographique (`amenity`, `shop`, `building`) tandis que le second indique l'état de cet objet. L'état dont il est question ici fait référence au cycle de vie de l'objet dont les états possibles sont définis sur le Wiki d'OSM¹⁷. Une église abandonnée pourra, par exemple, être décrite par le tag `amenity:abandoned=place_of_worship`.

16. https://www.loc.gov/standards/iso639-2/php/code_list.php

17. http://wiki.openstreetmap.org/wiki/Lifecycle_prefix

5.2.4 Représentation des différents types de valeur

Les valeurs (`of4osm:IOSMTagValue`) adjointes aux clés précédemment décrites présentent elles-aussi des formes variées que nous avons choisi d'organiser dans 6 types de valeur (partie droite de la figure 5.3). Chaque type de valeur correspond à un type de données primitif. Les 6 types de valeur sont les suivants :

Valeur chaîne de caractères (`of4osm:IOSMTagStringValue`) Les valeurs chaîne de caractères sont représentées sous la forme d'un seul lemme. Ce sont les valeurs les plus fréquentes dans la base de données OSM. Certaines valeurs correspondent à une catégorie ("`school`", "`restaurant`") de l'objet géographique auquel elles sont associées, d'autres à ses propriétés (son nom : "`Londres`", son créateur : "`user123@mailbox.com`" ou encore sa source : "`http://wiki.openstreetmap.org/`").

Valeur booléenne (`of4osm:IOSMTagBooleanValue`) Les valeurs booléennes sont représentées par les chaînes de caractères "`yes`" ou bien "`no`". Ce sont des valeurs associées à des clés qui ont valeur de propriétés telles que `internet_access=yes`.

Valeur numérique (`of4osm:IOSMTagNumericValue`) Les valeurs numériques représentent un gradient continu dans l'intensité d'une propriété. Par exemple, la hauteur d'un bâtiment peut être indiquée avec le tag `height=80`, son élévation par rapport au niveau de la mer avec le tag `elev=350`.

Valeur de type date (`of4osm:IOSMTagDateValue`) Les valeurs de type date représentent une marque chronologique ponctuelle. Par exemple, les tags `start_date=2000-01-01` et `end_date=2000-12-31` indiquent que l'objet auquel ils sont rattachés a existé pendant l'année 2000. Le format préconisé pour la description de la date est la norme ISO 8601¹⁸. Les tags composés de ce type de valeur représentent donc des propriétés.

Valeur domaines temporels (`of4osm:IOSMTagTimeDomainsValue`) Les valeurs domaines temporels expriment une plage horaire, un ensemble d'intervalles de temps. Les spécifications concernant la syntaxe de ce type de valeur sont données de façon exhaustive sur le Wiki d'OSM¹⁹). Ce type de valeur est essentiellement associé à la clé `opening_hours`, laquelle est associée à plus de 500 000 objets enregistrés dans la base de données OSM²⁰.

Valeur multiple (`of4osm:IOSMTagMultipleValue`) Les valeurs multiples consistent

18. <http://www.iso.org/iso/home/standards/iso8601.htm>

19. http://wiki.openstreetmap.org/wiki/Key:opening_hours/specification implémentée dans le convertisseur JavaScript `opening_hours.js` (https://github.com/opening-hours/opening_hours.js).

20. http://taginfo.openstreetmap.org/keys/opening_hours

en plusieurs lemmes séparés par un point-virgule ";"²¹. L'ordre des lemmes n'est pas significatif. Par conséquent, un tag possédant *n* valeurs peut être éclaté en *n* tags ayant la même clé et différentes valeurs. Par exemple, sur le plan sémantique, le tag `amenity=library;cafe` équivaut à l'ensemble de tags `amenity=library` et `amenity=cafe`. L'utilisation de cette forme de valeur est déconseillée sur le wiki d'OSM²². Cependant, on trouve un nombre significatif de tags de cette forme : certains objets géolocalisés proposent plusieurs services sur une même empreinte géographique (`amenity=parking;fuel`, `amenity=bar;restaurant`); d'autres sont fréquemment disposés à proximité (`amenity=bench;waste_basket`), la valeur multiple traduisant alors un manque de précision de la cartographie. L'annexe C donne la liste des valeurs multiples associées à la clé `amenity`, retrouvées grâce à l'API TagInfo²³.

5.2.5 Typologie des tags OSM

À partir de la combinaison des différents types de clé et de valeur précédemment définis, nous avons élaboré une typologie des tags OSM possibles. Cette typologie permet de caractériser chaque tag (`of4osm:IOSMTag`) présent dans la base de données OSM, afin de procéder, par la suite, à l'extraction d'une sémantique différente selon la catégorie à laquelle il appartient (cf. section 5.3). Ces différents types de tag ne correspondent pas au simple produit de tous les types de clé avec tous les types de valeur : certaines combinaisons sont incohérentes comme, par exemple, la combinaison d'une clé internationalisée (`of4osm:IOSMTagLocalizedKey`) avec une valeur booléenne (`of4osm:IOSMTagBooleanValue`). C'est la raison pour laquelle l'élaboration de la typologie exposée dans le tableau 5.1 est indispensable. Ce tableau recense et illustre les combinaisons possibles entre types de clé (colonne CLÉ) et de valeur (colonne VALEUR) pour former un type de tag défini dans notre typologie (colonne TYPE) qui est illustré par un exemple (colonne EXEMPLE).

Les tags de type `of4osm:IOSMSimpleKeyBooleanValuePropertyTag` et `of4osm:IOSMComplexKeyBooleanValuePropertyTag` ont tous deux une valeur de type booléenne (`of4osm:IOSMTagBooleanValue`) factorisée dans la super-classe `of4osm:IOSMBooleanValuePropertyTag`. Les tags de types `of4osm:IOSMSimpleKeyStringValuePropertyTag`, `of4osm:IOSMComplexKeyStringValuePropertyTag` et `of4osm:`

21. L'utilisation du point-virgule comme séparateur de valeurs complexes n'est pas consensuel et fait l'objet d'une discussion sur le Wiki OSM à l'adresse http://wiki.openstreetmap.org/wiki/Talk:Semi-colon_value_separator

22. http://wiki.openstreetmap.org/wiki/Semi-colon_value_separator

23. http://taginfo.openstreetmap.org/api/4/key/values?key=amenity&page=1&rp=999&sortname=count_all&sortorder=desc

	TYPE	CLÉ	VALEUR	EXEMPLE
<i>CategoryTag</i>	IOSMSimpleCategoryTag	IOSMTagsSimpleKey	IOSMTagsStringValue	amenity=school
	IOSMMultipleCategoryTag	IOSMTagsSimpleKey	IOSMTagsMultipleValue	amenity=cafe;library
	IOSMSStatefulCategoryTag	IOSMTagsStatefulKey	IOSMTagsStringValue	amenity:abandoned=theater
<i>PropertyTag</i>	IOSMSimpleKeyValuePropertyTag	IOSMTagsSimpleKey	IOSMTagsStringValue	name=London
	IOSMComplexKeyValueStringPropertyTag	IOSMTagsComplexKey	IOSMTagsStringValue	addr:street=JeanJures
	IOSMLocalizedKeyValueStringPropertyTag	IOSMTagsLocalizedKey	IOSMTagsStringValue	name:fr=London
	IOSMSimpleKeyValueBooleanPropertyTag	IOSMTagsSimpleKey	IOSMTagsBooleanValue	drive_in=false
	IOSMComplexKeyValueBooleanPropertyTag	IOSMTagsComplexKey	IOSMTagsBooleanValue	toilets:wheelchairs=true
	IOSMSimpleKeyValueNumericPropertyTag	IOSMTagsSimpleKey	IOSMTagsNumericValue	height=80
	IOSMComplexKeyValueNumericPropertyTag	IOSMTagsComplexKey	IOSMTagsNumericValue	addr:housenumber=123
	IOSMSimpleKeyValueDatePropertyTag	IOSMTagsSimpleKey	IOSMTagsDateValue	start_date=2000-01-01
	IOSMComplexKeyValueDatePropertyTag	IOSMTagsComplexKey	IOSMTagsDateValue	source:date=2000-01-01
	IOSMTimedDomainsValuePropertyTag	IOSMTagsSimpleKey	IOSMTagsTimeDomainsValue	opening_hours=Mo-Fr08:30-16:30
	IOSMMultipleValuePropertyTag	IOSMTagsSimpleKey	IOSMTagsMultipleValue	service=repair;drive-through

TABLE 5.1 – Typologie des tags OSM.

:IOSMLocalizedKeyStringValuePropertyTag possèdent une valeur de type chaîne de caractères (of4osm:IOSMTagStringValue) factorisée par la classe of4osm:IOSMStringValuePropertyTag. Les tags de types of4osm:IOSMSimpleKeyNumericValuePropertyTag et of4osm:IOSMComplexKeyNumericValuePropertyTag possèdent une valeur de type numérique (of4osm:IOSMTagNumericValue) factorisée par la classe of4osm:IOSMNumericValuePropertyTag. Les tags de types of4osm:IOSMSimpleKeyDateValuePropertyTag et of4osm:IOSMComplexKeyDateValuePropertyTag possèdent une valeur de type date (of4osm:IOSMTagDateValue) factorisée par la classe of4osm:IOSMDateValuePropertyTag. La figure 5.4 représente l'ensemble de la typologie décrite ci-dessus.

Cette typologie des tags OSM s'appuie sur la caractérisation de leur composants vus dans les sections 5.2.3 et 5.2.4. Cependant, les tags de type of4osm:IOSMSimpleCategoryTag et of4osm:IOSMSimpleKeyStringValuePropertyTag sont tous deux de la forme `<of4osm:IOSMTagSimpleKey >=<of4osm:IOSMTagStringValue >`. Afin de faire la distinction entre les tags regroupés dans l'un et l'autre de ces 2 types, le guide de bonnes pratiques qu'est le Wiki d'OSM est ici exploité. Les tags dont la clé est répertoriée dans la rubrique *Additional properties* sont de type of4osm:IOSMSimpleKeyStringValuePropertyTag, les autres de type of4osm:IOSMSimpleCategoryTag.

5.3 Une ontologie OSM plus expressive

Afin de générer une ontologie de concepts extraits des tags OSM dont la représentation est détaillée en section 5.2, le méta-modèle de OF4OSM dispose d'un schéma conceptuel en trois volets : 1) les concepts of4osm:IOSMCategoryTagKeyConcept et of4osm:IOSMCategoryTagConcept permettant de réutiliser la hiérarchie clé-valeur (cf. sous-section 5.3.1); 2) le concept of4osm:HighLevelConcept pour l'expression de concepts de haut niveau factorisant des caractéristiques de tags conceptuellement proches (cf. sous-section 5.3.2); 3) le concept of4osm:IAristotelicianConcept pour la représentation de la conjonction de plusieurs tags associés à une même instance (cf. sous-section 5.3.3). Chacun de ces concepts possède des attributs généraux factorisés dans le concept abstrait of4osm:IConcept. La figure 5.5 donne une représentation de ce modèle que nous détaillons dans les sections suivantes.

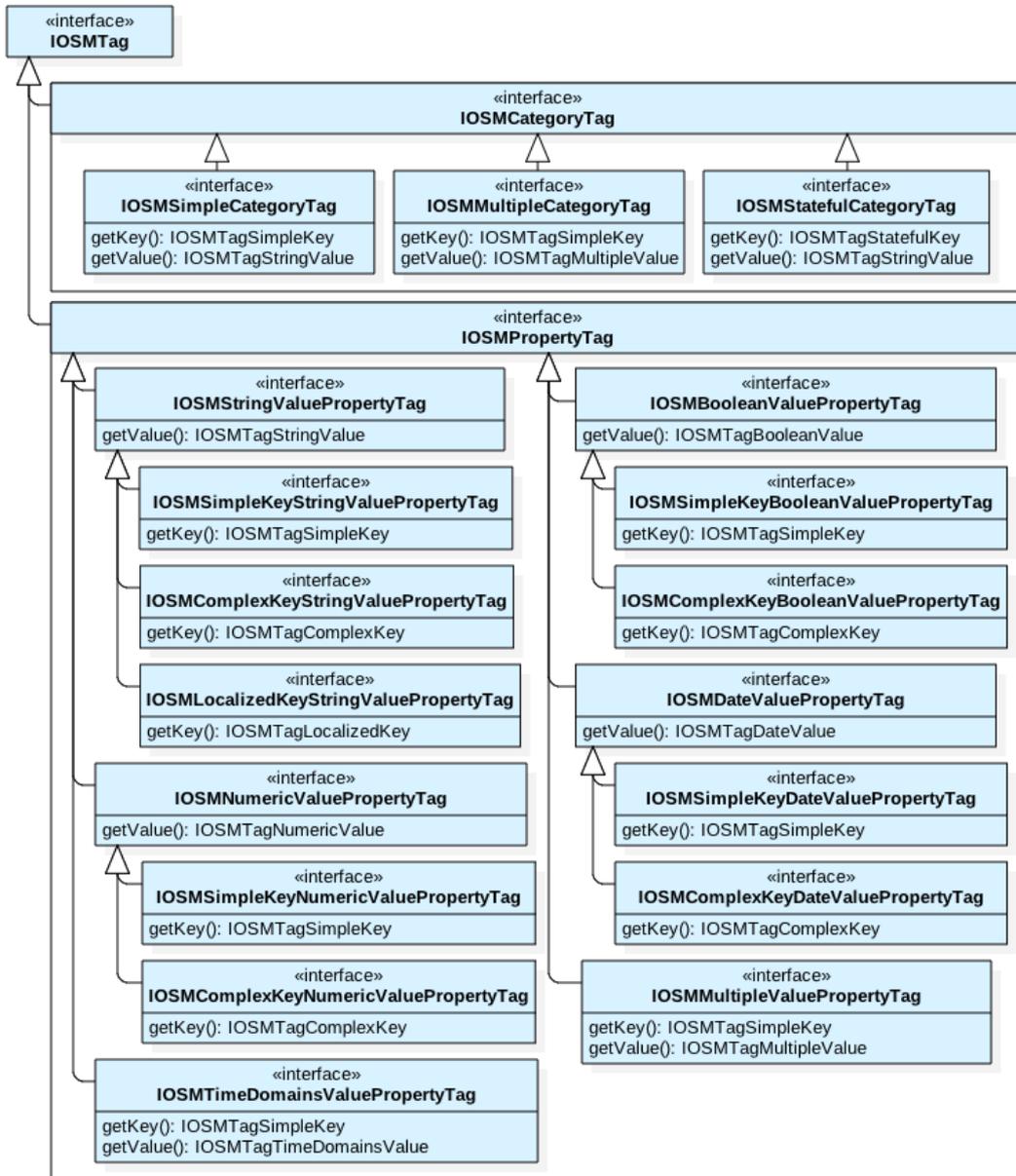


FIGURE 5.4 – Typologie des tags OSM.

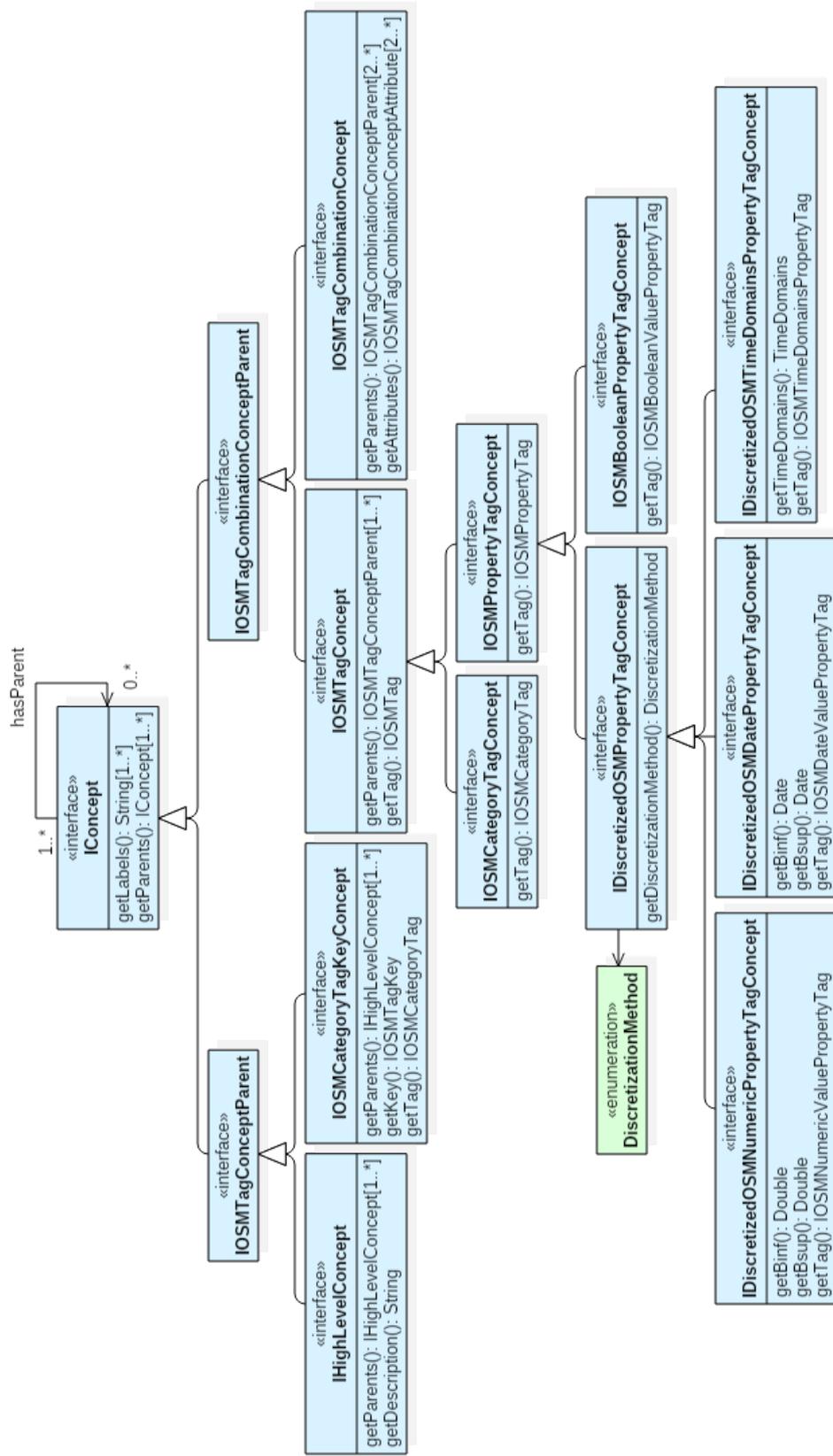


FIGURE 5.5 – Modèle de représentation des concepts extraits des tags OSM.

5.3.1 Hiérarchie initiale clé-tag

L'un des principaux intérêts de la représentation de connaissances au moyen d'ontologies est l'expression des relations entre les concepts. La plus élémentaire d'entre elles est la relation de subsomption. Malheureusement, les folksonomies sont intrinsèquement plates : il n'existe pas *a priori* de relations hiérarchiques entre les tags. En conséquence, les différentes tentatives de sémantisation des tags OSM présentées dans le chapitre 3 ont produit des taxonomies superficielles dont la profondeur est la résultante du seul rapport de subsomption entre clé et valeur qui, par ailleurs, n'a pas toujours de sens selon le type de tag.

La génération automatique de la hiérarchie clé-tag est le point de départ dans le processus de construction d'une ontologie extraite des tags OSM basée sur OF4OSM. Afin d'obtenir une classification pertinente, cette méthode nécessite de ne tenir compte que des tags qui représentent une catégorie, et non ceux qui ont valeur de propriété. Dans la lignée des travaux de [Auer et al., 2009], un lien de subsomption est donc établi entre chacun des tags de type catégorie (`of4osm:IOSMCategoryTag` : `of4osm:IOSMSimpleCategoryTag`, `of4osm:IOSMStatefulKeyCategoryTag` et `of4osm:IOSMMultipleCategoryTag`) et leur clé.

La taxonomie à deux niveaux ainsi obtenue est la même que celle issue de la méthode des auteurs de [Auer et al., 2009], à la différence près que nous considérons le tag complet comme sous-classe de sa clé, et plus seulement sa valeur. Plus précisément, nous introduisons deux concepts, `of4osm:IOSMCategoryTagKeyConcept` et `of4osm:IOSMCategoryTagConcept`, qui font respectivement référence au concept extrait de la clé et au concept extrait du tag complet pour les tags de type catégorie. En effet, il nous semble indispensable de conserver l'information de la clé dans le concept feuille de la hiérarchie clé-tag, car il existe des tags tels que `shop=bicycle` dont la valeur n'exprime pas à elle seule la sémantique du tag complet : le concept "vélo" (`bicycle`) n'est pas suffisant pour faire passer l'idée du "magasin de vélo" véhiculé par le tag `shop=bicycle`. Par conséquent, pour nommer le concept `of4osm:IOSMCategoryTagConcept`, les labels de la valeur et de la clé du tag associé sont concaténés sous la forme *ValeurClé*. Par exemple, à partir du tag `shop=bicycle` le concept `BicycleShop` est généré dans l'ontologie OF4OSM.

Ainsi, à partir d'un tag de type `of4osm:IOSMSimpleCategoryTag`, la hiérarchie initiale clé-tag est construite selon les trois étapes suivantes :

1. création d'un concept `of4osm:IOSMCategoryTagKeyConcept` nommé d'après la clé du tag (`of4osm:IOSMTagSimpleKey`). Ce concept est, ensuite, associé à la clé par la relation `skos:relatedTo` ;

2. création d'un concept `of4osm:IOSMCategoryTagConcept` nommé d'après le tag (concaténation de la valeur et de la clé au format *Camel Case*). Ce concept est, ensuite, associé au tag par la relation `skos:relatedTo` ;
3. création d'un lien de subsomption (`rdfs:subClassOf`) entre le concept de type `of4osm:IOSMCategoryTagKeyConcept` créé à l'étape 1) et le concept de type `of4osm:IOSMCategoryTagConcept` créé à l'étape 2).

Par exemple, à partir du tag `amenity=school` de type `of4osm:IOSMSimpleCategoryTag`, le concept `Amenity` de type `of4osm:IOSMCategoryTagKeyConcept` est créé puis associé, via la propriété `skos:relatedTo`, à la clé `of4osm:IOSMTagSimpleKey` correspondante. Ensuite ce concept est lié par une relation de subsomption (`rdfs:subClassOf`) au concept `SchoolAmenity` de type `of4osm:IOSMCategoryTagKeyConcept`. Ce dernier est ensuite associé (`skos:relatedTo`) au tag `of4osm:IOSMSimpleCategoryTag`. La figure 5.6 illustre le résultat de cette hiérarchisation.

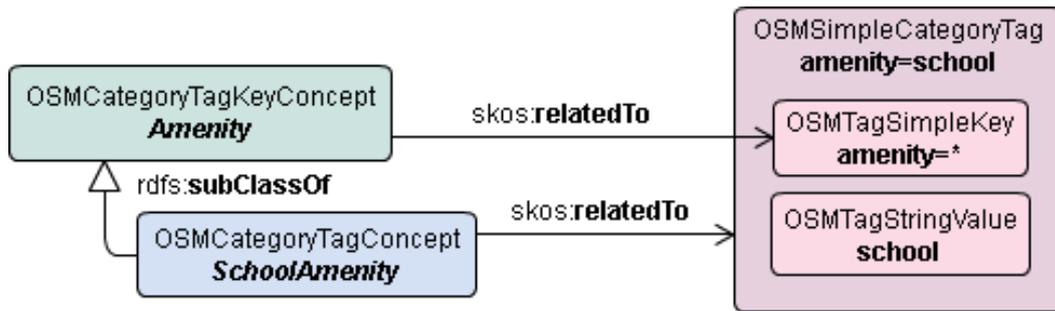


FIGURE 5.6 – Hiérarchie de concepts obtenue à partir du tag `amenity=school`

Il en va de même pour les tags de type `of4osm:IOSMStatefulKeyCategoryTag`. En effet, l'attribut relatif au statut du tag ne sera pas pris en considération puisqu'il est spécifique à l'instance et n'apporte rien sur le plan conceptuel. Ainsi, le tag `amenity:abandoned=place_of_worship` donnera lieu à la création des concepts `Amenity` (`of4osm:IOSMCategoryTagKeyConcept`) et `PlaceOfWorshipAmenity` (`of4osm:IOSMCategoryTagConcept`), tous deux liés par un lien de subsomption et respectivement lié à la clé (`of4osm:IOSMStatefulKeyCategoryTag`) du tag et au tag lui-même par la propriété `skos:relatedTo`.

À partir d'un tag de type `of4osm:IOSMMultipleCategoryTag` comme `amenity=library;cafe`, un concept de type `of4osm:IOSMCategoryTagKeyConcept` est créé à partir de la clé (`Amenity`) et associé à elle par la relation `skos:relatedTo`. Puis, autant de concepts de type `of4osm:IOSMCategoryTagConcept` qu'il y a de lemmes dans la valeur multiple (`LibraryAmenity` et `CafeAmenity`) sont créés, chacun de ces concepts étant, ensuite, associés au tag lui-même.

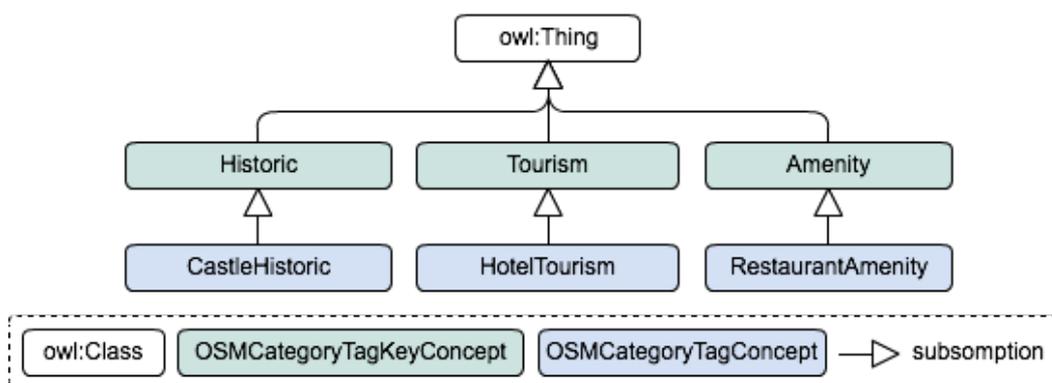


FIGURE 5.7 – Exemple de taxonomie générée à partir de la méthode de hiérarchisation clé-tag sur les tags associés au Château de la Commanderie.

Prenons l'exemple concret du Château de la Commanderie situé sur la commune d'Eybens dans le département de l'Isère. Les tags qui sont associés à cet objet sont les suivants : `historic=castle`, `tourism=hotel` et `amenity=restaurant`. En effet, il s'agit d'un château qui présente un intérêt historique en soi, mais il offre également la possibilité de s'y restaurer et d'y passer la nuit. La méthode de hiérarchisation clé-tag permet d'obtenir la taxonomie en figure 5.7.

5.3.2 Représentation de concepts de haut niveau

Pour parvenir à une hiérarchie plus profonde, nous proposons un concept `of4osm:IHighLevelConcept` qui n'est lié à aucune instance dans la base de données OSM, c'est-à-dire qu'il ne fait référence à aucune instance du concept `of4osm:IOSMTag`. `of4osm:IHighLevelConcept` sert à offrir la possibilité aux contributeurs OSM de créer une classe qui n'existe pas directement dans la base de données OSM, mais qui permet de factoriser, via une relation de subsumption (`rdfs:subClassOf`), des caractéristiques communes à des tags dont les clés sont différentes. Ainsi, des liens sémantiques peuvent être ajoutés entre des tags qui n'en avaient pas à l'issue de la génération de la taxonomie clé-tag (cf. section 5.3.1), mais entre lesquels il existe pourtant, d'après l'expertise de l'utilisateur OSM, une certaine similarité. Par conséquent, un `of4osm:IOSMTagConcept` est sous-classe d'un `of4osm:IOSMTagKeyConcept`, mais peut également être sous-classe d'un ou plusieurs `of4osm:IHighLevelConcept`, formant non plus une simple taxonomie, mais un Graphe Orienté Acyclique (*Acyclic Oriented Graph*, DAG) qui implémente le mécanisme d'héritage multiple.

Pour illustrer l'intérêt de ce modèle, prenons ici l'exemple d'un contributeur

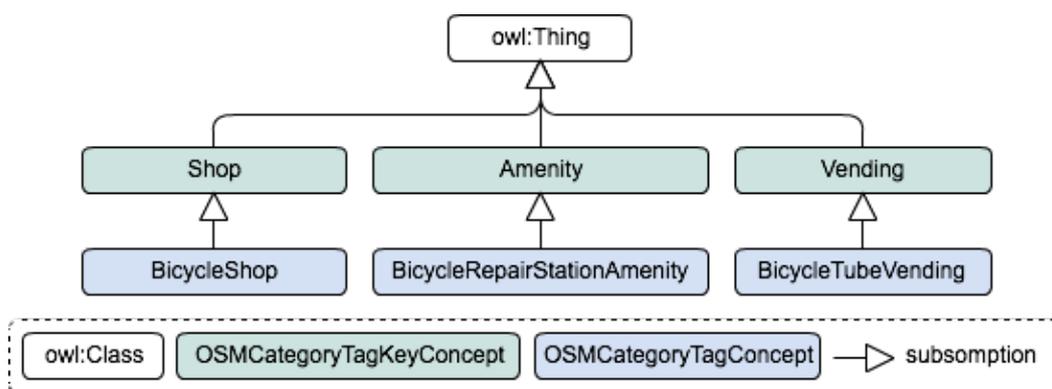


FIGURE 5.8 – Exemple de taxonomie liée au vélo générée sur la base de la hiérarchie clé-tag.

OSM qui voudrait associer à un distributeur de chambres à air pour vélo le tag `shop=bicycle`. Ce tag est très vague par rapport à l'objet qu'il caractérise. En effet, il existe un tag plus précis pour ce type d'objet : le tag `vending=bicycle_tube`. Si un système de suggestion de tags raisonne sur le réseau sémantique OSN dans lequel seules des relations de subsumption clé-valeur sont représentées (le concept `shop` est parent du concept `bicycle`), seuls les concepts qui ont également pour parent le concept `shop` seront retournés et proposés comme tags possibles à notre utilisateur. D'après le service Web TagInfo²⁴, les tags les plus utilisés ayant pour clé `shop` sont `shop=bakery`, `shop=clothes` et `shop=hairdresser`, lesquels présentent un intérêt pour le moins limité dans ce cas de figure. En dépit de sa pertinence, le tag `vending=bicycle_tube` ne sera pas retourné puisqu'il ne partage pas la clé `shop` avec le tag `shop=bicycle`.

Pourtant, les tags dont la sémantique est proche de `shop=bicycle` (`amenity=bicycle_repair_station` et `vending=bicycle_tube`) sont documentés sur le wiki d'OSM, ce qui indique que leur pertinence est communément admise. Chacun d'eux permet de décrire des équipements relatifs au vélo. En générant une taxonomie sur la base de la subsumption clé-tag, on obtient la hiérarchie représentée en figure 5.8. On constate donc que, bien que sémantiquement proches, aucun lien n'existe entre ces différents concepts. Par conséquent, sur la seule base d'une hiérarchie clé-tag, `shop=bicycle` est plus proche de `shop=hairdresser` (puisque'ils partagent la clé `shop`) que de `vending=bicycle_tube`. Ceci n'est pas satisfaisant dans le cadre d'une exploitation de l'ontologie pour un système de suggestion de tags lors de l'acquisition des données OSM.

Afin de factoriser le caractère commun à ces trois tags, un concept abstrait -

24. <http://taginfo.openstreetmap.org/keys/shop#values>

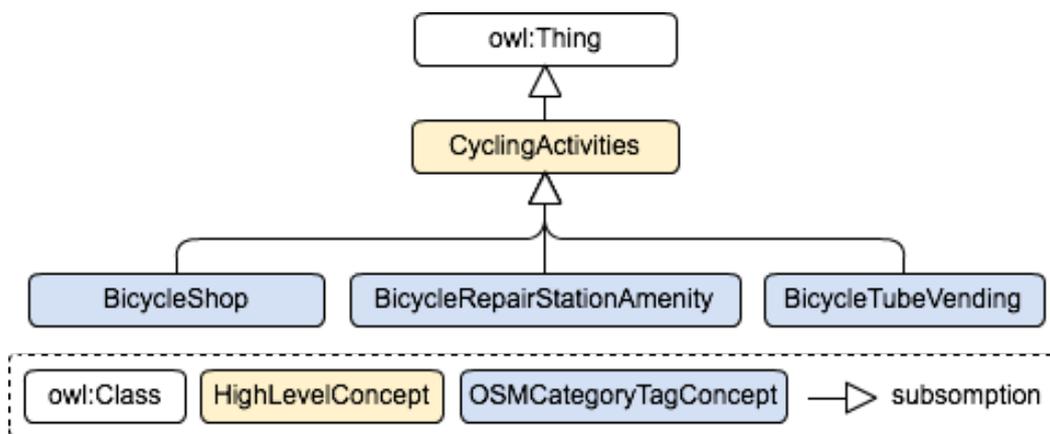


FIGURE 5.9 – Exemple de taxonomie liée au vélo construite grâce à l’instance - `CyclingActivities` du concept `of4osm:HighLevelConcept` pour lier des tags ayant des caractéristiques communes mais différentes clés.

`CyclingActivities`, instance du concept `of4osm:IHighLevelConcept` de OF4OSM, peut être créé par un contributeur OSM. La figure 5.9 présente une ontologie rudimentaire basée sur le méta-modèle OF4OSM, dans laquelle le concept abstrait `CyclingActivities` représente les activités liées au vélo. Dès lors, on peut rattacher les tags `shop=bicycle`, `amenity=bicycle_repair_station` et `vending=bicycle_tube` au concept `CyclingActivities`.

Avec le même système de suggestion basé sur un algorithme de similarité de structure, cette nouvelle classification permet de marquer les tags `amenity=bicycle_repair_station` et `vending=bicycle_tube` comme sémantiquement proches du tag `shop=bicycle` puisqu’ils partagent le parent `CyclingActivities`, et ce malgré le fait qu’ils n’aient pas de clé en commun. L’utilisateur est cette fois à même d’affiner sa description du distributeur de chambres à air. Même si, à ce stade, notre méta-modèle ne décrit que des relations hiérarchiques entre les tags, il permet de générer une taxonomie des tags issus de la folksonomie OSM plus expressive que celles étudiées dans le chapitre 3. La figure 5.10 donne une illustration de cette taxonomie pour les activités liées au cyclisme.

De façon plus générale, cette représentation permet à un utilisateur de retrouver des tags à partir de concepts de haut niveau qui factorisent un aspect sémantique commun à plusieurs tags. Le guide de bonnes pratiques d’OSM atteste implicitement de ce besoin : sur la page du Wiki OSM Map Features²⁵, les tags dont l’usage est prescrit sont catégorisés selon une taxonomie à trois niveaux. Le premier niveau, fait la distinction entre les tags utilisés pour décrire la nature des objets spatialisés dans

25. http://wiki.openstreetmap.org/wiki/Map_Features

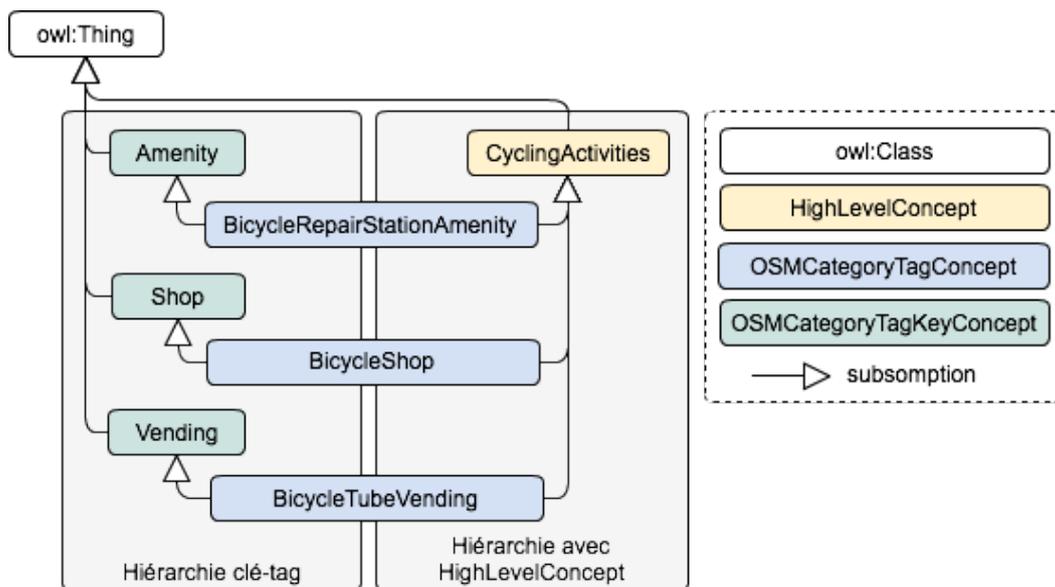


FIGURE 5.10 – Exemple de taxonomie liée au vélo construite grâce à la subsomption clé-tag et au concept haut-niveau `CyclingActivities`.

OSM (*Primary Features*²⁶, par exemple, `amenity=restaurant`, `craft=jeweller`, `shop=bicycle`, *etc.*) et les tags qui décrivent les propriétés de ces mêmes objets (*Additional Properties*²⁷, par exemple, `name=Tour Eiffel`, `height=300`, `addr:city=Paris`, *etc.*). Sous la rubrique *Primary Features*, un deuxième niveau regroupe des catégories dans lesquelles sont rangés les tags. Ces catégories correspondent aux clés des tags qu'elles recouvrent (*Amenity*, *Leisure*, *Shop*, *etc.*). Enfin, un troisième niveau permet de ranger plus finement ces tags selon des catégories qui, elles, ne correspondent à rien dans la forme du tag. Pourtant, ces catégories représentent bien une caractéristique commune entre plusieurs tags qui partagent des clés distinctes. À titre d'exemple, la catégorie de rang 2 *Amenity* regroupe les tags de type `amenity=*`. Ces derniers sont répartis dans les sous-catégories de rang 3 *Sustenance*, *Education*, *Transportation*, *Financial*, *Healthcare* et *Entertainment, Arts & Culture*. Chacune de ces dernières pourrait donc faire l'objet d'un concept de haut niveau `of-4osm:IHighLevelConcept` dans l'ontologie OF4OSM. Par conséquent, les concepts de haut niveau issus du guide de bonnes pratiques OSM seront ajoutés à l'ontologie OF4OSM au moment de son initialisation.

26. http://wiki.openstreetmap.org/wiki/Map_Features#Primary_features

27. http://wiki.openstreetmap.org/wiki/Map_Features#Additional_properties

5.3.3 Représentation de concepts aristotéliens

En plus de la méthode de subsomption clé-tag, OF4OSM dispose désormais d'un concept de haut niveau permettant à un contributeur OSM de factoriser des caractéristiques communes à des tags. Cependant, le niveau le plus fin dans cette ontologie est représenté par les tags : les instances de `of4osm:IHighLevelConcept` subsument éventuellement les clés `of4osm:IOSMTagKeyConcept`, lesquelles subsument à leur tour les tags `of4osm:IOSMTagConcept`. Afin d'obtenir une taxonomie profonde grâce à des concepts plus précis, nous exploitons l'utilisation conjointe de tags grâce à une méthode de classification automatique basée sur un modèle aristotélien.

5.3.3.1 Classification aristotélienne : principe général

Dans le domaine de la représentation des connaissances, pour procéder à la classification de concepts, l'approche ascendante est utilisée dans de nombreux domaines : les propriétés des entités sont *implicitement* analysées par des experts, leurs valeurs comparées puis les entités auxquelles elles appartiennent sont classifiées. Cependant, le caractère implicite de ce raisonnement est un frein majeur à ce travail d'expertise, précisément parce que l'analyse des propriétés qui justifient la classification n'est pas conservée. Un travail de rétro-ingénierie fastidieux est nécessaire lorsqu'il s'agit de reprendre l'édition de la classification. Cette entreprise est d'autant plus laborieuse si elle incombe à un expert qui n'en est pas l'auteur initial. Afin de rendre *explicite* cet exercice, nous avons collaboré au développement d'une ontologie pour la représentation de classes aristotéliennes multidimensionnelles.

Dans [Poole et al., 2008], les auteurs rappellent la définition aristotélienne d'une classe A : "Un A est un B tel que C" où B est la super-classe directe de A et C est le caractère qui rend A plus spécifique que B. Selon Aristote, B est le *genus* et C est le *differentia* [Sowa, 2000]. En collaboration avec les auteurs de [Poole et al., 2008], nous avons développé une représentation qui supporte ce point de vue ontologique. L'Ontologie Aristotélienne (*Aristotelian Ontology*, AO) que nous avons produite a pour but d'appuyer la description des classes par leurs propriétés (également appelées *dimensions*) afin de générer automatiquement une taxonomie de ces classes.

Dans cette ontologie, initialement développée pour un éditeur de classes aristotéliennes (*Aristotelian Class Editor*²⁸), une classe aristotélienne, `ao:IEntityClass`, possède une liste d'attributs (ou *differentia*), `ao:IAttribute`, eux-mêmes

28. <http://www.similar2.com:8080/ACE-Editor/>

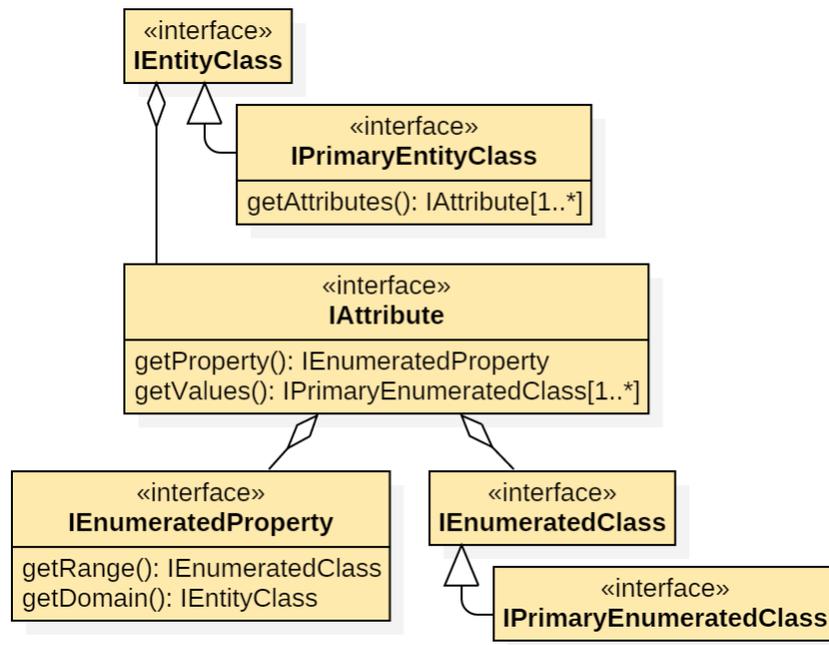


FIGURE 5.11 – Modèle simplifié de représentation d’une classe aristotélienne.

constitués de deux éléments : d’une part une propriété d’objet (`owl:ObjectProperty`), `ao:IEnumeratedProperty`, d’autre part la valeur de cette propriété, `ao:IEnumeratedClass`. La figure 5.11 est une version simplifiée de l’Ontologie Aristotélienne disponible en annexe D. Les concepts `ao:IPrimaryEntityClass` et `ao:IPrimaryEnumeratedClass` sont respectivement fils de `ao:IEntityClass` et `ao:IEnumeratedClass`, mais n’apportent pas de spécificités susceptibles d’enrichir le modèle OF4OSM. En effet, ils existent à l’origine dans l’AO dans le but d’exprimer la relation de synonymie : ils se différencient de leur concept frère, respectivement `ao:ISynonymEntityClass` et `ao:ISynonymEnumeratedClass`. Pour plus de clarté, ces derniers n’apparaissent pas sur la figure 5.11 mais ils sont présents dans l’annexe D représentant l’ontologie AO dans son intégralité.

5.3.3.2 La conjonction d’instances de `of4osm:IOSMTagConcept`

Si l’on considère les objets spatialisés dans la base de données OSM, on constate que nombre d’entre eux possèdent de multiples tags. Si l’on postule que chacun des tags qui les caractérisent représente une de leurs propriétés, alors on peut assimiler un tag à un *differentia*. Dès lors, si à une instance de la base de données OSM sont associés trois tags A, B et C, alors il existe une classe `A_B_C` (à laquelle appartient l’instance considérée) qui est sous-classe des classes `A_B` et `B_C`, elles-mêmes sous-classes des classes A, B et C. Afin de tirer parti de cette méthode, OF4OSM utilise

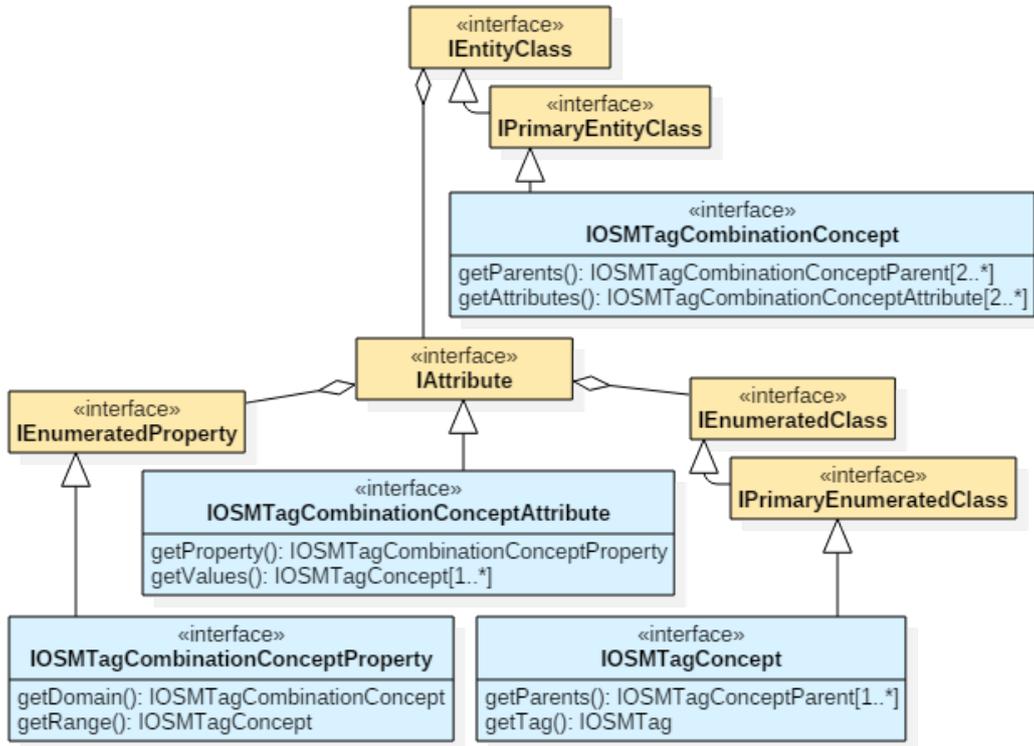


FIGURE 5.12 – Conformité du méta-modèle OF4OSM avec l’ontologie AO pour la représentation aristotélicienne de la conjonction de tags OSM.

le concept `of4osm:IOSMTagCombinationConcept` qui hérite du concept général `of4osm:IConcept` (cf. figure 5.5), mais également du concept `ao:IEntityClass` dans le modèle aristotélicien de l’ontologie AO.

Les *differentia* du concept `of4osm:IOSMTagCombinationConcept` sont en réalité les différents tags associés à une même instance, représentés dans notre modèle conceptuel par `of4osm:IOSMTagConcept` sur la figure 5.5. Afin de se conformer au modèle de l’ontologie AO présenté dans la figure 5.11, chaque `of4osm:IOSMTagConcept` est composé d’une liste d’attributs, `of4osm:IOSMTagCombinationAttribute`, sous-concept de `ao:IAttribute`. Chacun de ces attributs est lui-même composé d’une propriété `of4osm:IOSMTagCombinationConceptProperty` (fille de la propriété `ao:IEnumeratedProperty`) et d’une valeur `of4osm:IOSMTagConcept` (fille du concept `ao:IEnumeratedClass`). La figure 5.12 illustre la conformité du méta-modèle OF4OSM au modèle aristotélicien élaboré dans l’ontologie AO.

Pour illustrer cette approche, reprenons l’exemple du Château de la Commanderie à Eybens en Isère décrit dans la base de données OSM par le biais des tags `historic=castle`, `tourism=hotel` et `amenity=restaurant`. Selon cette description,

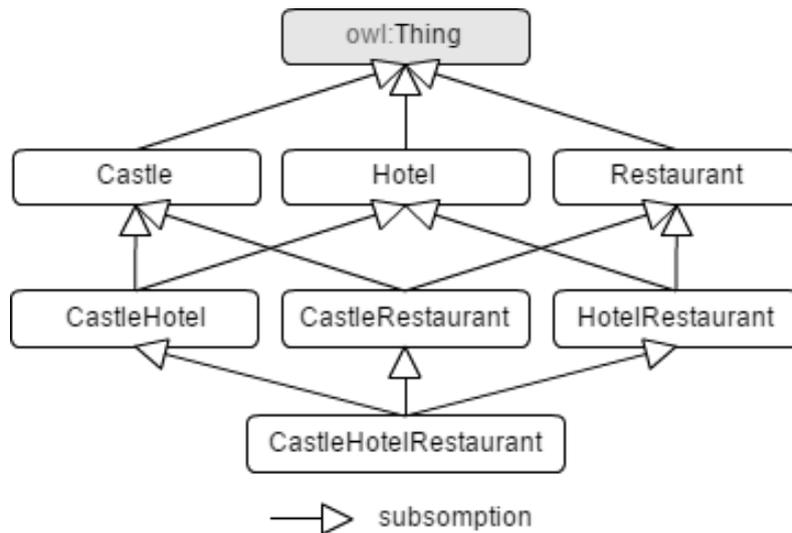


FIGURE 5.13 – Exemple de taxonomie générée à partir de la méthode aristotélicienne appliquée à la conjonction de tags OSM.

ce bâtiment est à la fois instance des classes Château, Hôtel et Restaurant. Or, il appartient aussi à la classe des objets étant à la fois Château et Hôtel, à celle des Hôtels-Restaurants, mais aussi à celle des Châteaux faisant office de Restaurant. La figure 5.13 donne la taxonomie générée 1) par la hiérarchisation clé-tag des tags associés au château, ainsi que 2) par la classification aristotélicienne sur les tags associés au Château. La classe `CastleHotelRestaurant` dont le château est une instance hérite à la fois des classes `Historic`, `Castle`, `Tourism`, `Hotel`, `Amenity` et `Restaurant` (hiérarchie clé-tag), mais également des classes `CastleHotel`, `CastleRestaurant`, `HotelRestaurant` (classification aristotélicienne).

L'ontologie ainsi obtenue est une hiérarchie à héritage multiple à 4 niveaux. Dans toutes les ontologies bâties sur le modèle aristotélicien, il existe une classe racine, la classe de plus haut niveau dans la taxonomie, mère de toutes les autres et qui ne possède pas de parent. Dans le formalisme du langage OWL, toute classe `owl:Class` est implicitement sous-classe de la classe `owl:Thing` : c'est donc cette classe qui servira de classe racine à l'ontologie extraite des tags OSM.

5.4 Conclusion

Au cours de ce chapitre, nous avons présenté en détail le méta-modèle qui vise à caractériser les tags de la folksonomie d'OSM afin de les structurer en une ontologie que nous appelons OF4OSM. Nous avons montré comment l'expressivité formelle

du méta-modèle de OF4OSM est supérieure à celle de la folksonomie OSM à partir de laquelle elle est extraite. Dès lors, cette ontologie permet, d'une part, d'améliorer la qualité des données OSM au moment de l'acquisition et, d'autre part, d'optimiser l'extraction des données OSM.

Dans un premier temps, en section 5.2, nous avons vu comment le méta-modèle supporte la caractérisation des tags en proposant une représentation spécifique des tags OSM. Cette représentation étend l'ontologie MUTO afin de proposer une représentation des tags OSM qui soit intégrée dans l'écosystème des vocabulaires pour la représentation des tags (cf. section 5.2.1). La conformité du méta-modèle de OF4OSM par rapport aux standards favorise l'interopérabilité des systèmes et donc la réutilisation des données qui en sont issues [Salzano, 2013]. La représentation de la structure des tags avec leurs composantes - la clé et la valeur - qui font la spécificité des tags OSM ont été présentées en sections 5.2.3 et 5.2.4. Selon le type de clé et de valeur dont ils sont constitués, les tags OSM correspondent à un certain type. La section 5.2.5 introduit la typologie des tags OSM proposée par le méta-modèle de OF4OSM. Cette caractérisation des tags permet d'extraire des concepts dont le rôle au sein de l'ontologie est différent afin de dégager une sémantique la plus précise possible.

Dans un second temps, la section 5.3 introduit le schéma conceptuel sur lequel est fondée l'ontologie OF4OSM extraite des tags OSM. Ce schéma supporte 3 approches distinctes : 1) une hiérarchisation à deux niveaux entre les concepts extraits des clés des tags et les concepts extraits des tags eux-mêmes (cf. sous-section 5.3.1), 2) la factorisation de caractéristiques partagées par plusieurs tags en des concepts de haut niveau (cf. sous-section 5.3.2), et 3) la création de concepts aristotéliens fondés sur la conjonction de différents tags associés à un même objet géographique (cf. section 5.3.3). La figure 5.14 illustre un exemple d'ontologie générée à partir du Château de la Commanderie à Eybens en Isère et qui implémente le méta-modèle de OF4OSM.



- historic = castle
- tourism = hotel
- amenity = restaurant

Le Château de la Commanderie à Eybens en Isère
<http://www.openstreetmap.org/way/66383517>

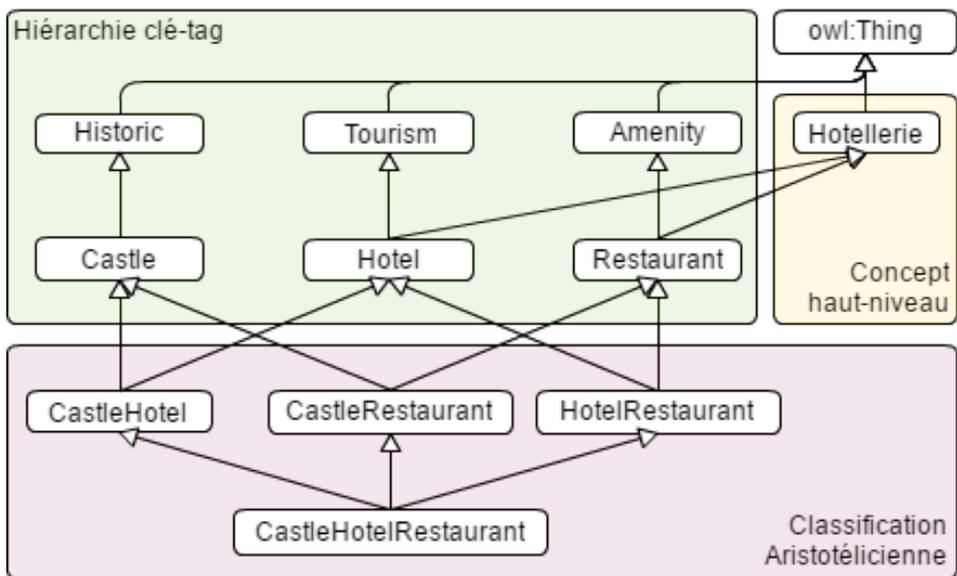


FIGURE 5.14 – Exemple d'ontologie générée à partir du méta-modèle de OF4OSM.

Chapitre 6

Génération de l'ontologie OF4OSM

6.1 Introduction

Dans le chapitre 5, nous avons exploré dans le détail les différents aspects du méta-modèle de l'Ontologie extraite de la Folksonomie pour OSM. Dans un premier temps, en section 5.2, la représentation des tags OpenStreetMap et de leur composantes, ainsi que la typologie qui en découle, ont été expliquées. Ceci permet de caractériser les tags issus de la base de données OSM. Dans un second temps, en section 5.3, un schéma conceptuel supportant 3 approches différentes a été exposé. La première approche consiste à générer automatiquement une hiérarchie à deux niveaux au moyen d'un lien de subsumption entre les concepts extraits de la clé et du tag qui contient cette clé, lorsque ce tag est de type *catégorie* (cf. section 5.3.1). La deuxième approche, semi-automatique, offre la possibilité à un contributeur d'ajouter des concepts de haut niveau qui ne soient pas directement extraits de la folksonomie d'OSM, mais qui permettent de factoriser une ou plusieurs caractéristiques partagées par des tags (cf. section 5.3.2). Enfin, la troisième approche vise à créer des classes aristotéliennes à partir de la conjonction des tags associés à un même objet géographique dans la base de données OSM (cf. section 5.3.3). Les deux parties du méta-modèle, 1) caractérisation et 2) structuration, permettent de supporter la génération de l'ontologie Ontologie extraite de la Folksonomie pour OSM.

Dans ce chapitre, nous nous intéressons à la génération de l'ontologie OF4OSM. Dans un premier temps, en section 6.2, nous présentons les méthodes et outils pour

l'alimentation initiale de l'ontologie OF4OSM à partir de la page *Map Features* du wiki OSM qui répertorie les tags consensuels acceptés par la communauté OSM. Dans un second temps, en section 6.3, nous introduisons les méthodes et outils pour l'alimentation complémentaire de l'ontologie OF4OSM au travers d'un processus découpé en 3 étapes. La première étape (section 6.3.1) consiste en l'extraction des données OSM via une API dédiée. Dans la deuxième étape (section 6.3.2), les tags sont caractérisés sur la base de la typologie des tags OSM vue en section 5.2.5. Enfin, dans la troisième partie (section 6.3.3), les méthodes de structuration des tags dans l'ontologie OF4OSM sont explicitées.

6.2 Alimentation initiale de l'ontologie OF4OSM à partir du wiki OSM

Comme évoqué en section 5.3.2, l'exploitation de la page *Map Features*¹ sur le wiki d'OSM est pertinente dans le cadre de notre approche. En effet, cette page Web répertorie des tags dont l'utilisation a d'ores et déjà fait l'objet d'un consensus au sein de la communauté OSM. Il nous semble donc judicieux d'initialiser l'ontologie OF4OSM sur la base des tags qu'elle contient et de la façon dont elle les structure. Dans cette section, la méthode de récupération du contenu de la page *Map features* est abordée en section 6.2.1, tandis que la méthode d'alimentation initiale de l'ontologie OF4OSM à partir de l'information récupérée sur cette page Web est expliquée en section 6.2.2

6.2.1 Récupération du contenu de la page *Map features*

Sur la page *Map features* du wiki OSM, les tags OSM sont rangés sous 3 niveaux de titres. Cette structure est illustrée sur la figure 6.1. Au premier niveau de la table des matières, on trouve les deux grands groupes de tags : 1) *Primary features* et 2) *Additional Property*. Ceux-ci correspondent respectivement aux tags de type *catégorie* (`of4osm:IOSMCategoryTag`) et aux tags de type *propriété* (`of4osm:IOSMPropertyTag`). Dans la lignée des travaux de [Auer et al., 2009], et comme nous l'avons vu dans le chapitre 5, nous ne nous intéressons qu'aux tags de type *catégorie*. Le second niveau de la hiérarchie de la table des matières de la page *Map Features* correspond aux clés des tags OSM. Par exemple, les titres 1.1) *Aerialway*, 1.2) *Aeroway* ou encore 1.3) *Amenity* visibles sur la figure 6.1 font référence à des clés utilisées dans des tags de la base de données OSM (par exemple, `aerialway=chair_-`

1. http://wiki.openstreetmap.org/wiki/Map_Features

1 Primary features		1 st level: category/property	
1.1 Aerialway		2 nd level: IOSMCategoryTagKey keys	
1.2 Aeroway			
1.3 Amenity			
1.3.1 Sustenance		3 rd level: IHighLevelConcepts concepts	
1.3.2 Education			
1.3.3 Transportation			
1.3.4 Financial			
1.3.5 Healthcare			
...			
content: IOSMCategoryTag tags			
Healthcare			
amenity	clinic		A medium-sized medical facility or health centre. 
amenity	dentist		A dentist practice / surgery. 
amenity	doctors		A doctor's practice / surgery. 

FIGURE 6.1 – Structure de la page *Map Features* du wiki OSM.

lift, aeroway=aerodrome ou encore amenity=school). Les tags de type *catégorie* qui partagent la même clé étant, *a priori*, plus proches sémantiquement que les autres tags, chaque concept de l'ontologie OF4OSM extrait d'un tag OSM sera lié à un concept représentant sa clé par une relation de subsomption (cf. section 5.3.1).

Afin de créer cette hiérarchie clé-tag, et puisque les titres de niveau 2 de la table des matières de la page Web *Map Features* correspondent aux clés des tags OSM, un concept `of4osm:IOSMCategoryTagKeyConcept` est créé pour chacun d'entre eux. Tous les concepts de l'ontologie OF4OSM disposent d'un attribut sous la forme d'un tableau associatif qui permet de stocker des labels internationalisés. Lors de la création d'un concept à partir d'une clé, le label correspondant au concept est la chaîne de caractères représentant la clé préalablement convertie au format *Upper Camel Case*². Par ailleurs, la page *Map Features* n'étant disponible qu'en anglais à ce jour, le label est associé au code langage "en_EN". Par exemple, à partir des

2. <http://c2.com/cgi/wiki?UpperCamelCase>

3 titres de niveau 2 1.1) *Aerialway*, 1.2) *Aeroway* et 1.3) *Amenity*, les 3 instances correspondantes, de type `of4osm:IOSMCategoryTagKeyConcept`, sont créées : les concepts *Aerialway*, *Aeroway* et *Amenity* sont alors ajoutés à l'ontologie OF4OSM.

Le niveau 3 de la table des matières est le niveau le plus fin sous lequel sont rangés les tags OSM sur la page *Map Features* du wiki OSM. Par exemple, sous le titre 1.3) *Amenity*, il existe les titres de niveau 3 1.3.1) *Sustenance*, 1.3.2) *Education*, 1.3.3) *Transportation*, etc. (cf figure 6.1). À la différence des titres de niveau 2 qui font référence aux clés des tags OSM, les titres de niveau 3 ne font pas référence à la base de données OSM. Par exemple, pour décrire les écoles dans OSM, il n'existe pas de tag `education=school` mais un tag `amenity=school`. Toutefois, cette classification implique que certains tags partagent une caractéristique commune implicite qui dépasse la sémantique exprimée par la clé. Parmi les tags qui partagent une même clé, il existe donc des groupes de tags entre lesquels le lien de similarité est plus étroit. Afin de tirer parti de cette sémantique, nous nous proposons d'ajouter le troisième niveau de la table des matières de la page *Map Features* à l'ontologie OF4OSM. Pour ce faire, un concept de haut niveau (`of4osm:IHighLevelConcept`) est créé pour chaque titre de niveau 3. Par exemple, le titre de niveau 1.3.2) *Education*, sous-niveau de 1.3) *Amenity*, lui-même sous-niveau de 1) *Primary features*, donne lieu à un concept de haut niveau auquel le label internationalisé "en_EN", "Education" est associé.

Le label est ici formé par la concaténation du titre de niveau 3 et du titre de niveau 2 sous lequel il se trouve, ceci dans le but de lever toute ambiguïté. En effet, la page *Map Features* peut potentiellement recenser des titres de niveau 3 identiques du point de vue de leur orthographe mais très différents sur le plan conceptuel. En l'état, la page *Map Features* ne compte qu'un seul cas d'ambiguïté : le titre de niveau 3 *Others*, placé sous les titres de niveau 2 *Amenity*, *Emergency* et *Shop*. Un traitement spécifique pour la chaîne de caractères "Others" peut être envisagé comme solution. Cependant, dans le but d'éviter que ce type de problème se pose dans des versions ultérieures du wiki OSM pour différents titres de niveau 3, le recours à la concaténation nous paraît plus pertinent pour lever toute ambiguïté.

Enfin, chacun des tags de la page *Map Features* du wiki OSM est utilisé pour créer un concept de type `of4osm:IOSMCategoryTagConcept` dans l'ontologie OF4OSM. Au moment de sa création, un label correspondant à la concaténation de sa valeur puis de sa clé, converti au format `UpperCamelCase`, lui est associé. À titre d'exemple, le tag `amenity=school` donne lieu au concept de type `of4osm:IOSMCategoryTagConcept` initialisé avec le label "en_EN", "SchoolAmenity". Ensuite, un lien de subsomption est établi entre le concept extrait d'un tag et les concepts

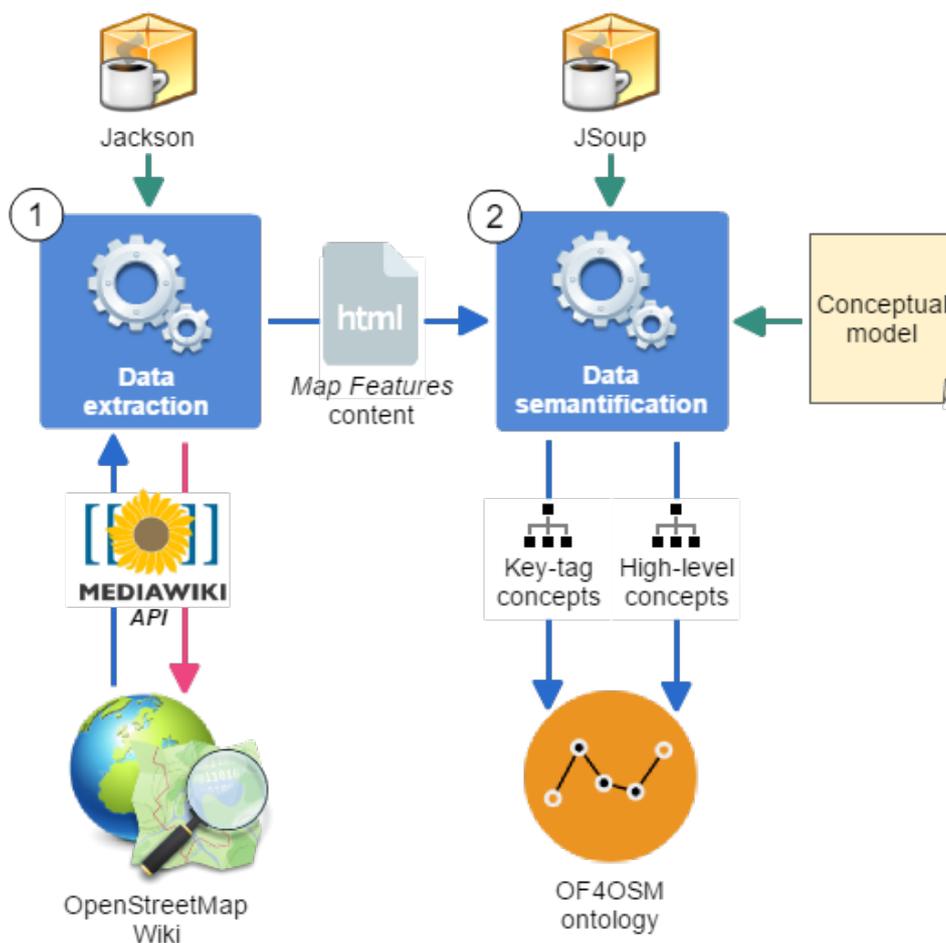


FIGURE 6.2 – Alimentation initiale de l’ontologie OF4OSM à partir de page *Map Features* du wiki OSM.

précédemment extraits des niveaux 2 et 3 de la hiérarchie de titres. Par exemple, le concept `SchoolAmenity` est lié au concept de type `IOSMCategoryTagKeyConceptAmenity` (extrait du titre de niveau 2 *1.3) Amenity*), mais également au concept `IHighLevelConceptEducation` (extrait du titre de niveau 3 *1.3.2) Education*).

Afin de pouvoir mettre à jour l’ontologie OF4OSM au gré des évolutions du wiki OSM, l’intégration des tags répertoriés sur la page *Map Features* se fait de façon automatique. La figure 6.2 illustre le processus en deux étapes que nous avons implémenté dans un prototype écrit dans le langage de programmation Java³. En l’état, notre prototype n’effectue pas de mises à jour en fonction des changements opérés sur le wiki. Cependant, le passage en revue des tags de la page *Map Features* peut être lancé périodiquement (par exemple, via une tâche cron⁴) afin de synchroniser

3. <https://www.java.com/>

4. <https://doc.ubuntu-fr.org/cron>

OF4OSM avec le wiki OSM.

La première étape opérée par le prototype consiste en l'extraction du contenu de la page *Map Features* du wiki OSM. Dans cette optique, nous avons réalisé un client HTTP qui interroge l'interface de programmation du système de gestion de contenu (*Content Management System*, CMS) sur lequel s'appuie le wiki OSM, à savoir MediaWiki⁵. Cette API permet de récupérer, pour une page donnée du wiki OSM, ses méta-données et son contenu textuel formaté en HTML (cf. annexe E). Le client HTTP fait appel à la bibliothèque Java Jackson⁶ dans le but de manipuler facilement les informations au format *JavaScript Object Notation* (JSON) renvoyées par le serveur.

6.2.2 Alimentation initiale de OF4OSM à partir du contenu de la page *Map features*

La seconde étape opérée par le prototype, et illustrée sur le schéma 6.2, consiste en l'analyse de l'information de la page *Map Features* extraite au cours de la première étape (cf. section 6.2.1). Le document HTML est alors manipulé au moyen de la bibliothèque Java JSoup⁷. Cette dernière offre un accès facilité au *Document Object Model* (DOM) du document HTML. Ainsi, le parcours des différents niveaux de titre de la page Web *Map Features* est effectué grâce à des sélecteurs *Cascading StyleSheet* (CSS)⁸. En s'appuyant sur le modèle conceptuel du méta-modèle de OF4OSM présenté en section 5.3, le système ajoute les concepts correspondant aux tags OSM (`of4osm:IOSMCategoryTagConcept`) répertoriés sur la page *Map Features*, ainsi que les concepts correspondant aux titres de niveau 2 (`of4osm:IOSMCategory-TagKeyConcept`) et 3 (`of4osm:IHighLevelConcept`) sous lesquels ces mêmes tags sont rangés. L'algorithme suivant décrit ce processus :

```
1 func initOF4OSM(of4osm : OF4OSMonto)
2 var mapFeaturesClient : MapFeaturesClient
3 var html : string
4 var htmlParser : HtmlParser
5 var htmlDoc : HtmlDocument
6 var primaryFeatures : HtmlElement
7 var siblings : array[HtmlElement]
8 var keyConcept : IOSMCategoryTagKeyConcept
9 var tagConcept : IOSMCategoryTagConcept
10 var highLevelConcept : IHighLevelConcept
11
```

5. http://www.mediawiki.org/wiki/API:Main_page

6. <https://github.com/FasterXML/jackson>

7. <https://jsoup.org/>

8. http://www.w3schools.com/cssref/css_selectors.asp

```

12 begin
13   html ← mapFeaturesClient.sendRequest()
14   htmlDoc ← parser.parse(html)
15   primaryFeatures ← htmlDoc.getElementById("Primary_features")
16   siblings ← primaryFeatures.getSiblings()
17   for each sibling in siblings do
18     if sibling.tagName = "h3" then
19       keyConcept ← createOSMCategoryTagKeyConcept(sibling.text)
20       addConcept(onto, keyConcept)
21     else if siblings.tagName = "h4" then
22       highLevelConcept ← createHighLevelConcept(sibling.text)
23       addParent(highLevelConcept, keyConcept)
24       addConcept(onto, highLevelConcept)
25     else if siblings.tagName = "li" then
26       tagConcept ← createOSMCategoryTagConcept(sibling.text)
27       addParent(tagConcept, highLevelConcept)
28       addConcept(onto, tagConcept)
29     end if
30   end for
31 end

```

Extrait de code 6.1 – Algorithme d’alimentation initiale de l’ontologie OF4OSM à partir de la page *Map Features* du wiki OSM.

L’algorithme 6.1 présente la fonction d’alimentation initiale de l’ontologie OF4OSM à partir de la page *Map Features* du wiki OSM. Cette fonction prend en paramètre d’entrée l’ontologie OF4OSM. Dans un premier temps, le client HTTP interroge l’API MediaWiki (ligne 13) dans le but de récupérer le contenu de la page HTML *Map Features*. Ensuite, le document HTML est converti en un objet (ligne 14) permettant d’accéder facilement à ses balises. Puis, le premier titre de niveau 1, représentant les tags de type *catégorie*, est atteint (ligne 15). Les éléments frères sont alors récupérés (ligne 16) parcourus un à un (lignes 17 à 28). Dans la boucle, si l’élément courant est un titre de niveau 2 (par exemple, *1.3 Amenity* en figure 6.1, associé ici à la balise `<h3>`, car la balise `h1` caractérise le titre de la page *Map Features*, et non les titres de niveau 1 dans la table des matières), un concept `of4osm:OSMCategoryTagKeyConcept` est créé et ajouté à l’ontologie (lignes 18 à 21). Si l’élément courant est un titre de niveau 3 (par exemple, *1.3. Education* en figure 6.1, associé ici à la balise `<h4>`), un concept `of4osm:HighLevelConcept` est créé, le dernier concept `of4osm:OSMCategoryTagKeyConcept` lui est ajouté comme parent, puis il est ajouté à l’ontologie (lignes 21 à 25). Si l’élément courant est un article de liste (par exemple, `amenity=clinic` en figure 6.1, associé ici à la balise ``), un concept `of4osm:OSMCategoryTagConcept` est créé, le dernier concept `of4osm:HighLevelConcept` lui est ajouté comme parent, puis il est ajouté à l’ontologie (lignes 25 à 29).

Le temps d’exécution du code 6.1 sur une machine dotée d’un processeur de 2.2Ghz et de 8Go de RAM est d’une dizaine de secondes. Cet algorithme consiste en

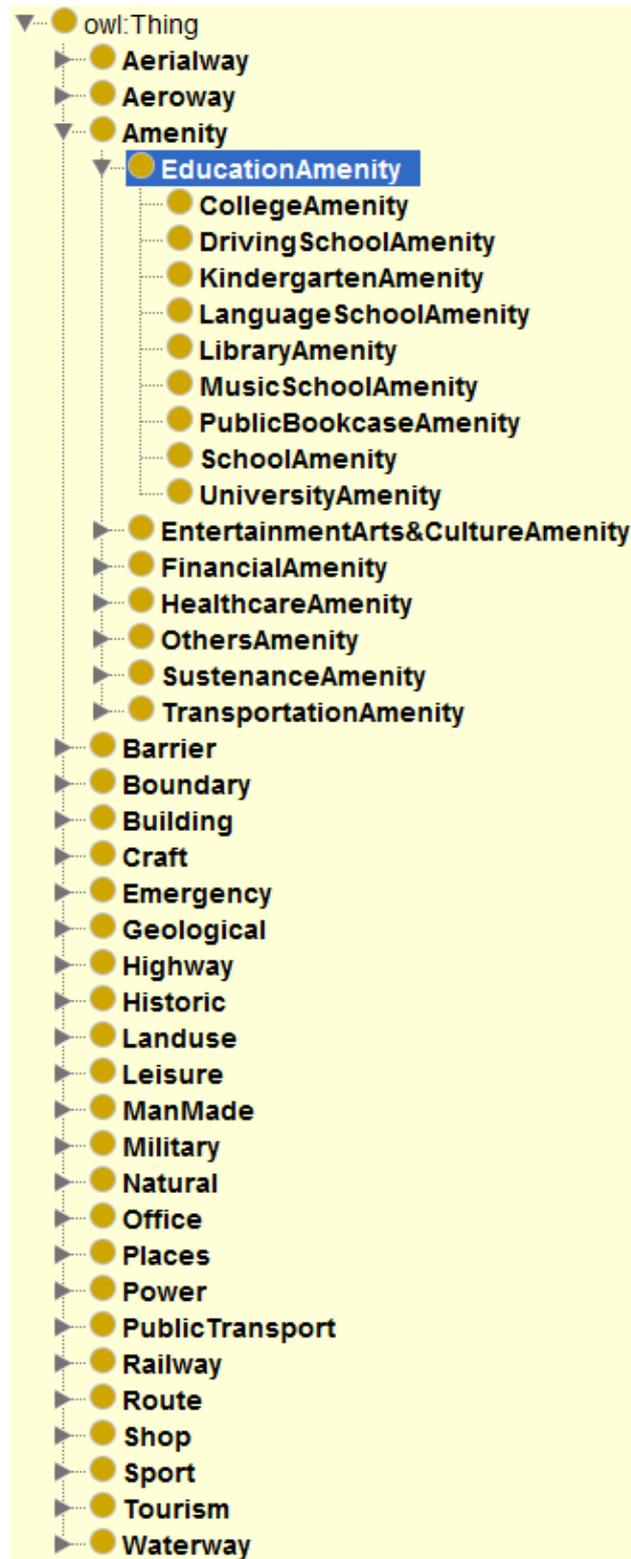


FIGURE 6.3 – Visualisation de l'ontologie OF4OSM après alimentation initiale dans l'éditeur Protégé.

une structure itérative dont la complexité est linéaire ($O(n)$). Nous avons réalisé l'alimentation initiale de l'ontologie OF4OSM sur la base de la version du 01/10/2016 de la page *Map Features* du wiki OSM. Le code du prototype qui implémente ce processus ainsi que l'ontologie qui en résulte sont disponibles à l'adresse suivante : <https://github.com/anthonyhombiat/OF4OSM>. La figure 6.3 illustre OF4OSM telle que visualisée au moyen de l'interface de navigation proposée par l'éditeur d'ontologies Protégé⁹. Après alimentation initiale, l'ontologie OF4OSM est composée de 1 038 concepts organisés en une taxonomie à 3 niveaux. La répartition de ces concepts est la suivante :

- 26 instances de `of4osm:IOSMCategoryTagKeyConcept` ;
- 58 instances de `of4osm:IHighLevelConcept` ;
- 898 instances de `of4osm:IOSMCategoryTagConcept`.

Une fois ces tags consensuels collectés à partir du guide de bonnes pratiques du projet OSM, nous pouvons encore enrichir l'ontologie OF4OSM avec les tags associés aux objets géographiques de la base de données OSM. La section 6.3 aborde ce processus d'alimentation complémentaire de l'ontologie OF4OSM.

6.3 Alimentation complémentaire de l'ontologie OF4OSM à partir de la base de données OSM

Dans cette section, nous introduisons les méthodes et outils pour peupler l'ontologie OF4OSM à partir de la base de données OSM. La figure 6.4 illustre les 3 étapes de ce processus. La première étape (section 6.3.1), consiste en l'extraction des données OSM au moyen d'une API dédiée et de la bibliothèque JAXB pour la manipulation des données OSM au format XML. La deuxième étape (section 6.3.2) sert à caractériser les tags OSM sur la base de la typologie exposée en section 5.2.5. La troisième et dernière partie (section 6.3.3) s'appuie sur le modèle conceptuel du méta-modèle de OF4OSM vu en section 5.3 pour ajouter des concepts issus des tags OSM à l'ontologie OF4OSM.

6.3.1 Récupération des tags OSM

Pour procéder à l'extraction de l'information en provenance de la base de données OSM, l'API OSM, brièvement présentée en section 2.4, permet de retrouver l'information située sur une zone géographique dont la surface n'excède pas 0.25 degrés

9. <http://protege.stanford.edu/>

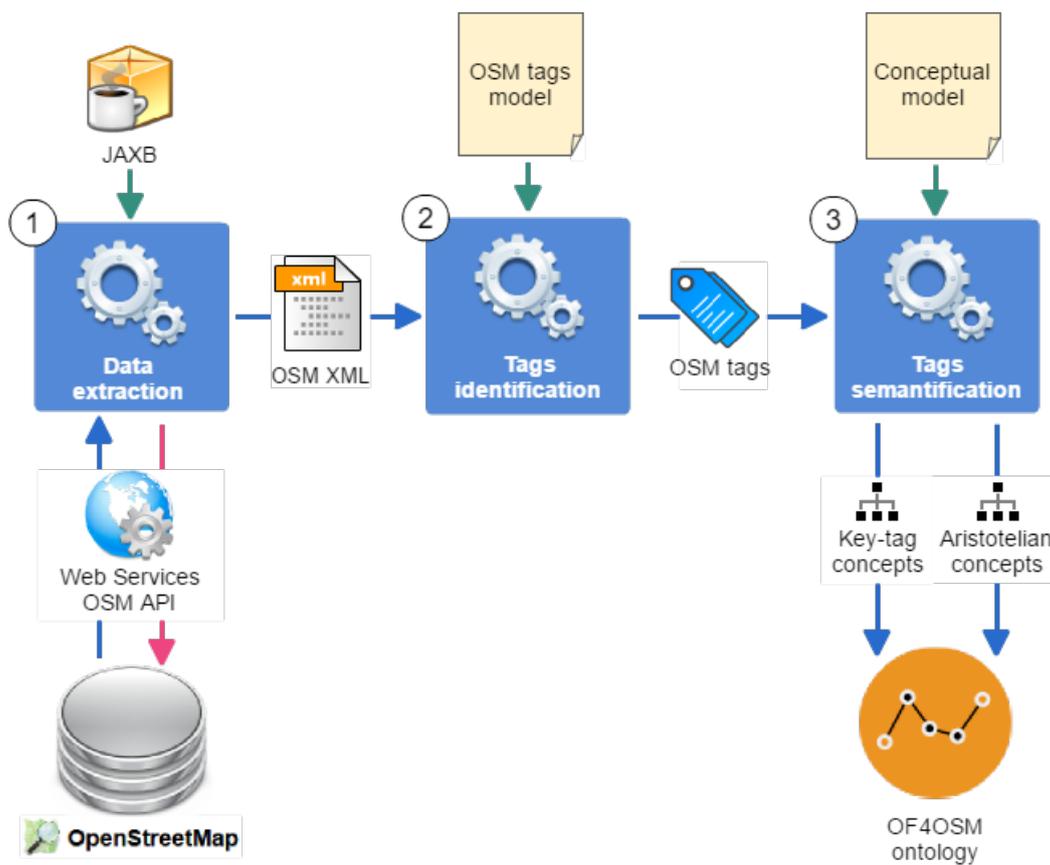


FIGURE 6.4 – Processus d’alimentation complémentaire de l’ontologie OF4OSM à partir des tags issus de la base de données OSM.

carrés et dont le nombre d’éléments à retourner est limité à 50 000. Afin de trouver un compromis entre la représentativité de notre aire d’étude et le temps d’exécution de nos expériences, nous avons choisi la ville de Grenoble (BBOX : 5.677606, 45.15414, 5.753118, 45.214077) comme aire d’étude¹⁰. Grenoble comporte, en date du 1er octobre 2016, 83 112 éléments cartographiés (noeuds, lignes et relations confondus) dans la base de données OSM. L’API OSM ne répondant pas à nos besoins en termes de quantité d’éléments à traiter (83 112 > 50 000), nous avons mis à contribution une autre interface de programmation : l’API OSM eXtended API (Xapi).

Xapi est une surcouche de l’API OSM¹¹ permettant d’interroger la base de données OSM en lecture seule. Elle offre la possibilité d’exécuter des requêtes sur une zone géographique dont la taille est circonscrite à 100 degrés carrés et le nombre

10. Pour obtenir la BBOX correspondant à la ville de Grenoble, nous avons utilisé le site Web <http://boundingbox.klokantech.com/>. Dans la suite de nos expériences, Grenoble fera référence à la zone délimitée par cette BBOX

11. http://wiki.openstreetmap.org/wiki/API_v0.6

d'objets limité à 10 millions¹². Cette interface possède une architecture de type REST qui fonctionne sur le protocole HTTP. Par conséquent, nous avons élaboré un client HTTP dédié à l'interrogation de l'interface de l'API Xapi. Le principal serveur conforme à cette API est disponible à l'adresse suivante : `http://www.overpass-api.de/api/xapi`. La réponse d'un serveur implémentant l'API Xapi - modélisée dans la figure 6.5 - est conforme à l'API d'OSM classique. Cette réponse est composée d'une collection de noeuds (**Node**), lignes **Way** et relations **Relation** - les 3 types de donnée primitifs de la base de données OSM (cf. section 2.4) - représentant les objets géographiques situés sur la BBOX (**Bounds**) passée en paramètre de requête par le client. Chacun de ces types primitifs possède un certain nombre d'attributs en commun (factorisés dans la classe abstraite **Element**), parmi lesquels une liste de tags (**Tag**).

Afin de montrer le fonctionnement de l'approche, et étant donné que la base de données OSM prise dans sa totalité représente plus de 666GB¹³, le périmètre de notre expérimentation est circonscrit à la commune de Grenoble. La requête suivante permet de retrouver tous les objets présents dans la base de données OSM qui sont situés sur la zone définie par la boîte englobante (*Bounding Box*, BBOX) correspondant à la ville de Grenoble :

```
1 GET http://www.overpass-api.de/api/xapi?
2   *[bbox=5.677606,45.15414,5.753118,45.214077]
```

De la même façon que l'API OSM de base, la surcouche Xapi utilise le format XML. Afin de traiter les réponses au format XML du serveur implémentant l'API OSM, la bibliothèque *Java Architecture for XML Binding* (JAXB)¹⁴ est utilisée. L'extrait de code 6.2 ci-dessous est un extrait de la réponse reçue par notre client :

```
1 ...
2 <node id="234" visible="true" version="2" changeset="234"
3   timestamp="2014-04-24T10:46:17Z" user="user38" uid="123"
4   lat="45.188" lon="5.717">
5   <tag k="name" v="Salon GHD"/>
6   <tag k="shop" v="hairdresser"/>
7   <tag k="source" v="survey"/>
8 </node>
9 <node id="456" visible="true" version="3" changeset="345"
10  timestamp="2014-04-24T10:46:18Z" user="user38" uid="123"
11  lat="45.188" lon="5.716">
12  <tag k="name" v="Tropical Food"/>
13  <tag k="shop" v="convenience"/>
14  <tag k="source" v="survey"/>
```

12. <http://wiki.openstreetmap.org/wiki/Xapi#Notes>

13. <http://wiki.openstreetmap.org/wiki/Planet.osm>

14. <https://jaxb.java.net/>

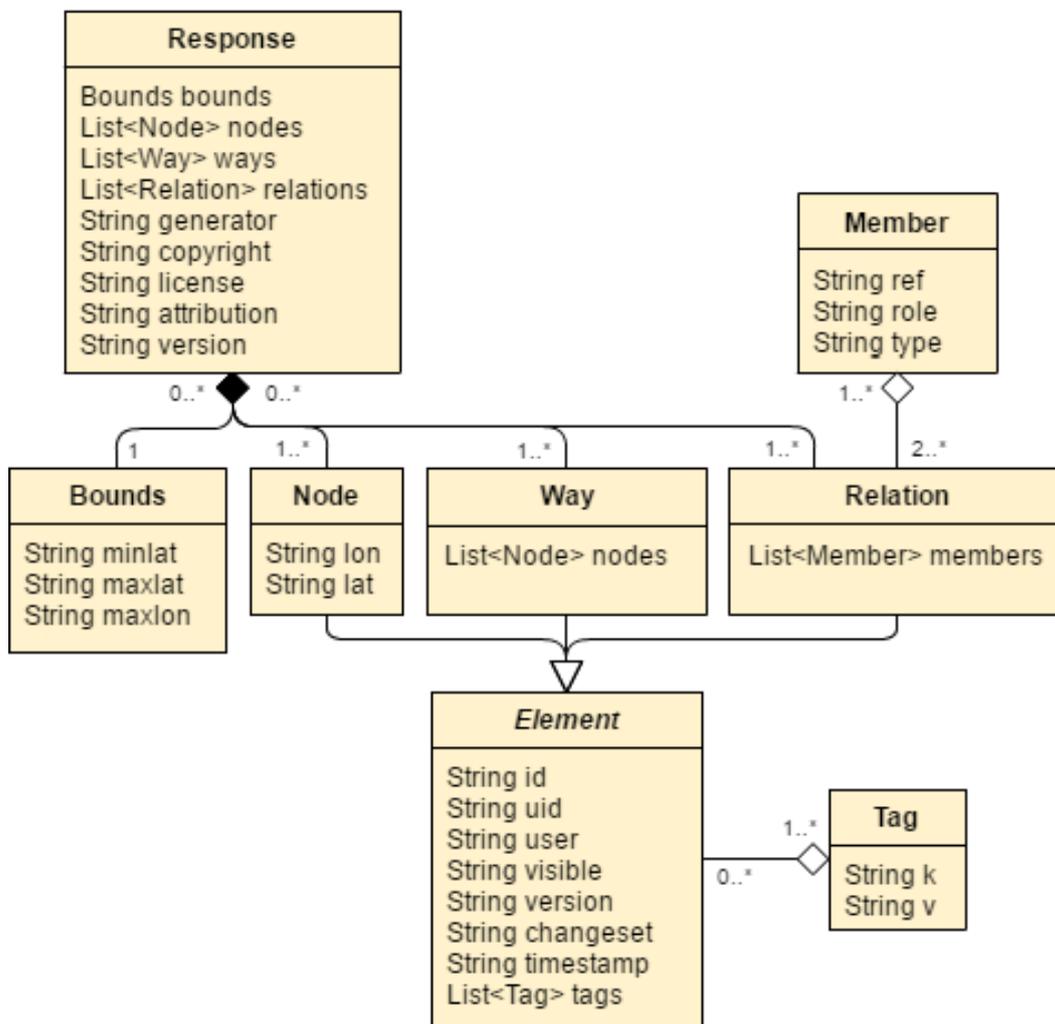


FIGURE 6.5 – Implémentation dans notre prototype de la réponse renvoyée par un serveur conforme à l'API OSM.

```

15 </node>
16 <node id="5" visible="true" version="4" changeset="456"
17   timestamp="2015-08-28T05:27:02Z" user="user38" uid="123"
18   lat="45.188" lon="5.716">
19   <tag k="amenity" v="kindergarten"/>
20   <tag k="name" v="Les Bambins d'Abord"/>
21   <tag k="source" v="survey"/>
22 </node>
23 ...

```

Extrait de code 6.2 – Extrait de la réponse du serveur implémentant l'API OSM représentant des objets géographiques situés sur la ville de Grenoble

Dans l'extrait 6.2, 3 objets géographiques sont représentés par des noeuds. Le premier (lignes 2 à 8) est un coiffeur (`shop=haidresser`), le second (lignes 9 à 15)

est une épicerie (`shop=convenience`), le troisième (lignes 16 à 22) est une garderie (`amenity=kindergarten`). À partir de cet ensemble d'objets et des tags qui leur sont associés, l'étape de caractérisation des tags peut être opérée.

6.3.2 Caractérisation des tags

Une fois les tags OSM collectés, nous procédons à leur caractérisation dans le but d'en extraire, *in fine*, une ontologie de concepts. Dans cette section, la méthode permettant de caractériser les tags et leurs composantes - la clé et la valeur - est expliquée.

6.3.2.1 Caractérisation des clés

Dans un premier temps, ce sont les clés des tags OSM collectés comme indiqué en section 6.3.1 qui sont caractérisées. Ce processus s'appuie sur le modèle conceptuel du méta-modèle de OF4OSM présenté en section 5.3. Plus précisément, il s'agit ici de mettre à contribution la représentation du tag OSM et de ses composantes, ici la clé, telle que décrit en section 5.2.3. Conformément à l'API OSM, le service Web implémentant l'API Xapi renvoie la clé d'un tag sous la forme d'une chaîne de caractères, valeur de l'attribut `k` de la balise `<tag>` (lignes 5 à 7, 12 à 14 et 19 à 21 sur l'extrait de code 6.2). Cette chaîne peut correspondre à 4 types de clés : 1) les clés simples, 2) les clés complexes, 3) les clés avec état et 4) les clés internationalisées.

Afin d'identifier auquel de ces 4 types correspond une clé donnée, une analyse de la chaîne de caractères correspondant à la clé du tag OSM est opérée. La fonction `identifyKey(key:string)` présentée dans l'algorithme 6.3 synthétise cette étape de caractérisation de la clé. Dans ce processus, la chaîne représentant la clé est d'abord éclatée en un tableau de chaîne de caractères sur la base du caractère séparateur ":" (ligne 6). En effet, ce dernier sépare les lemmes d'une clé complexe, avec état ou encore internationalisée (cf. section 5.2.3). Si le tableau ne contient qu'une seule chaîne de caractères alors il s'agit d'une clé simple `of4osm:IOSMTagSimpleKey` (ligne 8). Sinon, si le tableau contient plus de deux chaînes de caractères, il s'agit d'une clé complexe `of4osm:IOSMTagComplexKey` (ligne 14). Sinon, si le tableau contient exactement deux chaînes de caractères et que la seconde correspond au format des codes de langages ISO-639¹⁵, il s'agit d'une clé internationalisée `of4osm:IOSMTag-LocalizedKey` (ligne 10). Sinon, si le tableau contient exactement deux chaînes de caractères et que la seconde correspond à un des différents préfixes représentant des

15. http://www.iso.org/iso/home/standards/language_codes.htm

états dans le cycle de vie d'un objet de la base de données OSM inventoriés sur le wiki¹⁶, il s'agit d'une clé avec état `of4osm:IOSMTagStatefulKey` (ligne 12).

```
1 func identifyKey(key : string) : IOSMTagKey
2 const languages : array[string]
3 const states : array[string]
4 var splitKey : array[string]
5 begin
6   splitKey ← split(key, ':')
7   if size(splitKey) = 1 then
8     return createOSMTagSimpleKey(key)
9   else if size(splitKey) = 2 and contains(languages, splitKey[2]) then
10    return createOSMTagLocalizedKey(key)
11  else if size(splitKey) = 2 and contains(states, splitKey[2]) then
12    return createOSMTagStatefulKey(key)
13  else
14    return createOSMTagComplexKey(key)
15  end if
16 end
```

Extrait de code 6.3 – Algorithme de caractérisation des clés des tags OSM.

Le temps d'exécution du code 6.3 sur une machine dotée d'un processeur de 2.2Ghz et de 8Go de RAM est inférieur à une seconde. Cet algorithme consiste en une structure conditionnelle dont la complexité est constante ($O(1)$).

6.3.2.2 Caractérisation des valeurs

Après avoir identifié le type de la clé du tag OSM, c'est au tour de la valeur de faire l'objet d'une caractérisation. Celle-ci est basée sur la partie du méta-modèle de OF4OSM qui s'intéresse à la représentation de la valeur d'un tag OSM introduite en section 5.2.4. Selon cette représentation, la valeur d'un tag OSM est déclinée en 6 types différents : 1) valeur chaîne de caractères, 2) valeur booléenne, 3) valeur numérique, 4) valeur de type date, 5) valeur domaines temporels et 6) valeur multiple. Pour ne considérer que la valeur du tag OSM fourni par le service Web sous la forme d'une chaîne de caractères au format `<clé>=<valeur>`, celle-ci est découpée pour ne retenir que la sous-chaine suivant le caractère '='.

Pour identifier lequel de ces 6 types correspond à une valeur donnée, une analyse de la chaîne de caractères correspondant à la valeur du tag OSM est opérée. La fonction `identifyValue(tagValue:string)` présentée dans l'algorithme 6.4 synthétise cette étape de caractérisation de la valeur. Pour cela, la chaîne est d'abord éclatée sur la base du caractère séparateur ";" qui sert à séparer les tag à valeurs

16. http://wiki.openstreetmap.org/wiki/Lifecycle_prefix

mutiples (ligne 4). Si la taille du tableau qui résulte de ce découpage est supérieure à 1, il s'agit d'une valeur multiple `of4osm:IOSMTagMultipleValue` (ligne 15). Dans le cas contraire, le type de la chaîne est comparé aux types booléen (ligne 6), numérique (ligne 8), date (ligne 10) et domaines temporels (ligne 12). Les domaines temporels sont un type dont nous obtenons la validation grâce à la bibliothèque JavaScript `opening_hours.js`¹⁷. Si l'un de ces types correspond à la chaîne, l'instance de valeur de tag `of4osm:IOSMTagValue` correspondante est créée. Si la chaîne est booléenne, une instance de `of4osm:OSMTagBooleanValue` est créée (ligne 7). Si la chaîne est numérique, une instance de `of4osm:OSMTagNumericValue` est créée (ligne 9). Si la chaîne est une date, une instance de `of4osm:OSMTagDateValue` est créée (ligne 11). Si la chaîne représente des domaines temporels, une instance de `of4osm:OSMTagTimeDomainsValue` est créée (ligne 13). Sinon, une instance de (`of4osm:OSMTagStringValue`) est créée (ligne 15).

```

1 func identifyValue(value : string) : IOSMTagValue
2 var splitValue : array[string]
3 begin
4   splitValue ← split(value, ';')
5   if size(splitValue) = 1 then
6     if typeOf(value) is Boolean then
7       return createOSMTagBooleanValue(value)
8     else if typeOf(value) is Numeric then
9       return createOSMTagNumericValue(value)
10    else if typeOf(value) is Date then
11      return createOSMTagDateValue(value)
12    else if typeOf(value) is TimeDomains then
13      return createOSMTagTimeDomainsValue(value)
14    else
15      return createOSMTagStringValue(value)
16    end if
17  else
18    return createOSMTagMultipleValue(value)
19  end if
20 end

```

Extrait de code 6.4 – Algorithme de caractérisation des valeurs des tags OSM.

Le temps d'exécution du code 6.4 sur une machine dotée d'un processeur de 2.2Ghz et de 8Go de RAM est inférieur à une seconde. Cet algorithme consiste en une structure conditionnelle dont la complexité est linéaire ($O(1)$).

17. https://github.com/opening-hours/opening_hours.js

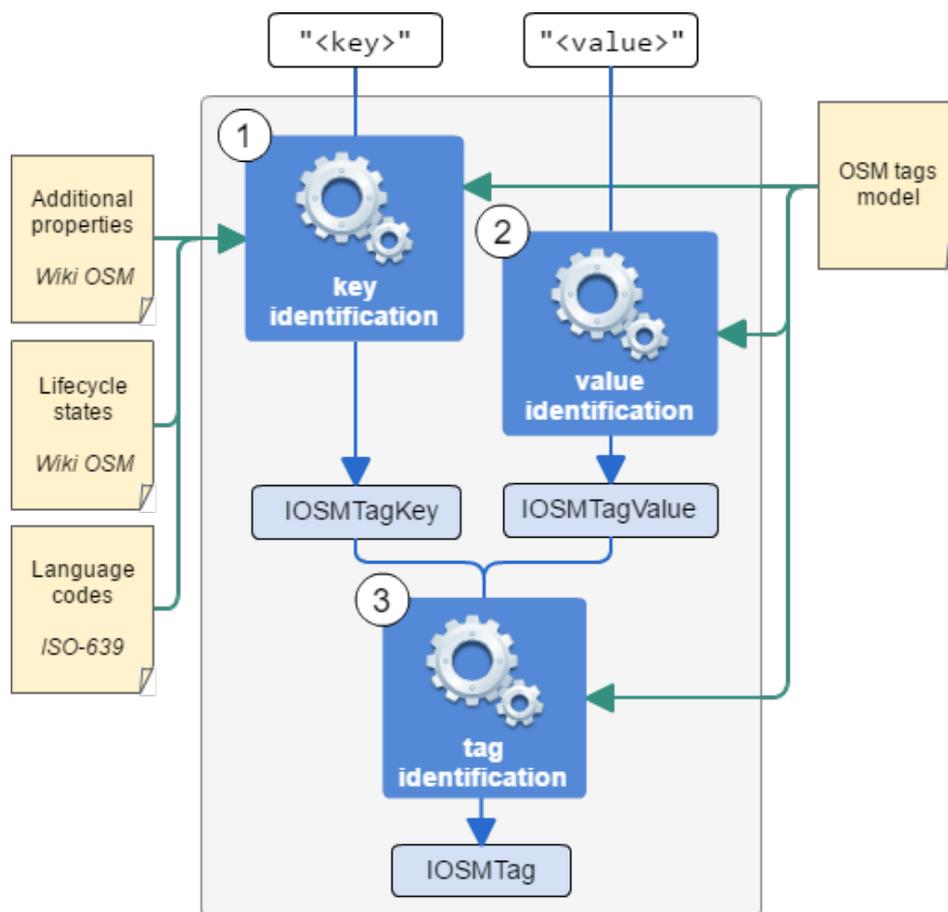


FIGURE 6.6 – Processus de caractérisation d'un tag OSM.

6.3.2.3 Caractérisation des tags

L'ensemble du processus de caractérisation d'un tag OSM est illustré dans la figure 6.6. Les étapes 1) de caractérisation de la clé (section 6.3.2.1) et 2) de caractérisation de la valeur (section 6.3.2.2) doivent être exécutées avant de pouvoir déterminer le type du tag OSM `of4osm:IOSMTag` (3^{ème} étape sur le schéma de la figure 6.6). La caractérisation est effectuée pour tous les tags associés à tous les objets géographiques situés sur la ville de Grenoble récupérés par le client Xapi (cf. section 6.3.1).

Pour procéder à la caractérisation d'un tag OSM, il faut, tout d'abord, discriminer les tags de type *catégorie* de ceux de type *propriété*. Ce tri dichotomique s'appuie sur la page *Map Features* du wiki OSM : une énumération contenant l'ensemble des clés répertoriées sous la rubrique *Additional Properties* est créée. Chaque clé de chaque tag est, par la suite, comparée à chacune des valeurs contenues dans cette énumération. Si la clé correspond à l'une de ces valeurs, alors il s'agit d'un tag de

type *propriété* (`of4osm:IOSMPropertyTag`). Sinon, il s’agit d’un tag de type *catégorie* (`of4osm:IOSMCategoryTag`). Par exemple, concernant le tag `amenity=school`, la clé `amenity` ne fait pas partie de l’énumération regroupant les clés des tags regroupés sous la rubrique *Additionnal Properties* : il s’agit donc d’un tag de type *catégorie* (`of4osm:IOSMCategoryTag`).

Ensuite, pour affiner le type de tag, la caractérisation se base sur la typologie des tags OSM du méta-modèle de OF4OSM présentée dans la section 5.2.5. Il s’agit ici de combiner les types de clé et de valeur identifiés aux étapes 1) et 2) de la figure 6.6 selon une structure conditionnelle qui, à un type de clé et un type de valeur, associe un type de tag. Si l’on considère de nouveau le tag `amenity=school`, sa clé est identifiée comme correspondant au type `of4osm:IOSMTagSimpleKey` et sa valeur comme correspondant au type `of4osm:IOSMTagStringValue`. À partir de ces deux types, et selon le tableau 5.1 associant à un type de tag OSM, un type de clé et un type de valeur, le type du tag `amenity=school` est déterminé : `of4osm:IOSMSimpleCategoryTag`.

TABLE 6.1 – Bilan quantitatif des tags OSM identifiés sur la ville de Grenoble

Type de tag OSM (<code>of4osm:IOSMTag</code>)	nb instances
<code>of4osm:IOSMSimpleCategoryTag</code>	47 091
<code>of4osm:IOSMStatefulCategoryTag</code>	80
<code>of4osm:IOSMMultipleCategoryTag</code>	130
<code>of4osm:IOSMSimpleKeyStringValuePropertyTag</code>	55 214
<code>of4osm:IOSMComplexKeyStringValuePropertyTag</code>	14 305
<code>of4osm:IOSMLocalizedKeyStringValuePropertyTag</code>	68
<code>of4osm:IOSMSimpleKeyBooleanValuePropertyTag</code>	45 126
<code>of4osm:IOSMComplexKeyBooleanValuePropertyTag</code>	995
<code>of4osm:IOSMSimpleKeyNumericValuePropertyTag</code>	7 712
<code>of4osm:IOSMComplexKeyNumericValuePropertyTag</code>	11 002
<code>of4osm:IOSMSimpleKeyDateValuePropertyTag</code>	0
<code>of4osm:IOSMComplexKeyDateValuePropertyTag</code>	3
<code>of4osm:IOSMTimeDomainsValuePropertyTag</code>	403
<code>of4osm:IOSMMultipleValuePropertyTag</code>	19 767
Total	201 970

Le temps d’exécution du processus de caractérisation des tags OSM associés aux objets géographiques situés sur la ville de Grenoble est de 10 minutes sur une machine dotée d’un processeur de 2.2Ghz et de 8Go de RAM. Le tableau 6.1 dresse le bilan quantitatif des 201 970 tags OSM associés sur cette zone géographique à

plus de 80 000 objets (en moyenne, sur cette zone, un objet est donc décrit par 2 à 3 tags). Les ensembles de tags présentés dans le tableau 6.1 comportent des doublons. En effet, dans la phase de classification aristotélicienne abordée en section 6.3.3.2, l'information reliant un tag donné à l'objet auquel il est associé est indispensable. Par conséquent, les tags doublons qui peuvent être associés à différents objets sont conservés.

6.3.3 Structuration des tags

À partir des tags caractérisés dans la section 6.3.2, la structure de l'ontologie OF4OSM peut être dégagée. Dans cette section, les principes de la structuration de la folksonomie d'OSM en l'ontologie OF4OSM sont présentés dans le détail. Ce processus implique 2 étapes distinctes : 1) la création d'une hiérarchie de concepts basée sur le postulat d'un lien de subsumption entre la clé d'un tag et le tag lui-même (section 6.3.3.1) et 2) la génération d'une hiérarchie de concepts extraits de la conjonction de tags associés à un même objet géographique (section 6.3.3.2). Le troisième axe de notre approche de structuration des tags OSM qui consiste en la factorisation de caractéristiques communes à certains tags en des concepts de haut niveau (`of4osm:IHighLevelConcept`, cf. section 5.3.2) à été mis à profit lors de l'alimentation initiale de l'ontologie OF4OSM à partir du wiki OSM expliquée en section 6.2.2.

6.3.3.1 Génération de la hiérarchie clé-tag

Les tags `of4osm:IOSMSimpleCategoryTag` caractérisés en section 6.3.2 sont ici exploités dans l'optique de générer la hiérarchie clé-tag supportée par le modèle conceptuel du méta-modèle de OF4OSM et détaillée en section 5.3.1. L'algorithme 6.5 formalise le fonctionnement de cette procédure. Dans un premier temps, les tags de type `of4osm:IOSMCategoryTag` (`of4osm:IOSMSimpleCategoryTag`, `of4osm:IOSMStatefulKeyCategoryTag` et `of4osm:IOSMMultipleCategoryTag`) sont parcourus (ligne 5). Pour chacun de ces tags, 2 concepts sont créés : un concept `of4osm:IOSMCategoryTagKeyConcept` (ligne 8) et un concept `of4osm:IOSMCategoryTagConcept` (ligne 9). Chacun d'eux possède un attribut sous la forme d'un tableau associatif permettant de stocker des labels internationalisés.

Lors de la création du premier concept, `of4osm:IOSMCategoryTagKeyConcept`, ce tableau associatif est initialisé avec la valeur correspondant à la clé du tag qu'on aura préalablement modifiée pour qu'elle soit conforme à la convention de nommage

Upper Camel Case. Par exemple, le tag `shop=bicycle` donne lieu à l'entrée "en_EN", "Shop" dans le tableau associatif. La langue préconisée par le wiki OSM est l'anglais, c'est pourquoi, lors de l'alimentation initiale, le code de langage `en_EN` est inséré dans le tableau. Par la suite, ces labels permettront aux utilisateurs d'ajouter les traductions des noms des concepts ainsi créés.

Lors de la création du second concept, `of4osm:IOSMCategoryTagConcept`, le tableau associatif permettant de stocker les labels est initialisé avec la valeur correspondant à la concaténation de la valeur puis de la clé du tag. Ainsi, le tag `shop=bicycle` donnera lieu au label "en_EN", "BicycleShop", ici aussi au format *Upper Camel Case*. Ce label sera également utilisé pour construire l'identifiant (IRI) du concept dans l'ontologie OF4OSM. Une fois encore, l'attribut contenant les labels donnera toute latitude au contributeur OSM pour adjoindre à la valeur initiale en anglais les traductions correspondantes.

Une fois les deux concepts, `of4osm:IOSMCategoryTagKeyConcept` et `of4osm:IOSMCategoryTagConcept`, créés, ils sont liés par une relation de subsomption (ligne 10). Enfin, chacun de ces deux concepts est ajouté à l'ontologie OF4OSM s'il ne l'est pas déjà (lignes 11 à 16), les doublons ne sont pas retenus.

```

1 func addKeyTagTaxo(
2   onto : OF4OSMOntology,
3   categoryTags : array [IOSMCategoryTag])
4 var keyConcept : IOSMCategoryTagKeyConcept
5 var valueConcept : IOSMCategoryTagConcept
6 begin
7   for each tag in categoryTags do
8     keyConcept ← createOSMCategoryTagKeyConcept(tag.key)
9     valueConcept ← createOSMCategoryTagConcept(tag)
10    addParent(valueConcept, keyConcept)
11    if !contains(onto, keyConcept) then
12      add(onto, keyConcept)
13    end if
14    if !contains(onto, valueConcept) then
15      add(onto, valueConcept)
16    end if
17  end for
18 end

```

Extrait de code 6.5 – Algorithme de création de la hiérarchie de concepts issue des tags OSM et de leur clé.

Pour les tags associés aux objets géographiques situés sur la ville de Grenoble, le temps d'exécution du code 6.5 sur une machine dotée d'un processeur de 2.2Ghz et de 8Go de RAM est d'une dizaine de secondes. Cet algorithme consiste en une structure itérative imbriquée, sa complexité est donc quadratique ($O(n^2)$), avec n le

nombre de tags. Après génération de la hiérarchie de concepts, les concepts suivants sont ajoutés à l'ontologie OF4OSM :

- 179 instances de `of4osm:IOSMCategoryTagKeyConcept` ;
- 1248 instances de `of4osm:IOSMCategoryTagConcept`.

À ce stade, en tenant compte des chiffres après alimentation initiale (cf. section 6.2.2), le bilan quantitatif des différents types de concepts de l'ontologie OF4OSM est donc de :

- 58 (58 + 0) instances de `of4osm:IHighLevelConcept` ;
- 205 (26 + 179) instances de `of4osm:IOSMCategoryTagKeyConcept` ;
- 2146 (898 + 1248) instances de `of4osm:IOSMCategoryTagConcept`.

Au terme de la structuration des tags OSM selon la hiérarchie clé-tag, l'ontologie OF4OSM compte près de sept fois plus de concepts liés aux clés des tags (`of4osm:IOSMCategoryTagKeyConcept`) et plus de 70% de concepts liés aux tags supplémentaires (`of4osm:IOSMCategoryTagConcept`). Cette augmentation significative du nombre de concepts témoigne de la portée englobante de notre approche qui vise à prendre en compte la grande diversité des tags effectivement utilisés par les contributeurs OSM, au-delà de l'ensemble restreint des tags documentés sur le wiki OSM.

6.3.3.2 Génération des classes aristotéliennes

La création de concepts aristotéliens `of4osm:IOSMTagCombinationConcept` consiste en la génération de classes extraites de la conjonction de tags de type `of4osm:IOSMCategoryTag` associés à une même instance (cf. section 5.3.3). Pour ajouter ces classes à l'ontologie OF4OSM, l'ensemble des objets géographiques collectés dans la base de données OSM (cf. section 6.3.1) est parcouru. Pour chacun de ces objets, toutes les combinaisons de tags sont passées en revue. L'ensemble des combinaisons pour une instance associée à n tags est ici réalisé par la permutation des n tags sans répétition et sans tenir compte de l'ordre : on parle d'arrangements. Le nombre d'arrangements A de k objets parmi n est défini par la formule suivante :

$$A_n^k = \frac{n!}{(n-k)!} \text{ pour } k \leq n$$

L'ensemble de ces arrangements est généré par l'algorithme récursif 6.6. Sa structure itérative lui confère une complexité linéaire ($O(n)$).

```

1 func combinations(
2   elements : array[Object],
3   length : int,
4   startPosition : int,
5   combination : array[Object],
6   result : array[array[Object]])
7 var i : int
8 begin
9   if length = 0 then
10     add(combination, result)
11   return
12   end if
13   i ← startPosition
14   for i <= size(elements) - length do
15     combination[size(combination) - length] ← elements[i]
16     combinations(elements, length-1, i+1, combination, result)
17     i ← i+1
18   end for
19 end

```

Extrait de code 6.6 – Algorithme pour la génération d'arrangements de k éléments parmi n.

Pour un ensemble de 4 concepts A, B, C, D , et avec $length=2$ (2 parmi 4) l'algorithme 6.6 renvoie les arrangements suivants :

$(A, B) (B, C) (C, D)$
 $(A, C) (B, D)$
 (A, D)

L'algorithme 6.7 réalise le traitement itératif qui exécute l'algorithme 6.6 (ligne 9) pour trouver les arrangements de 2 à n éléments parmi n éléments. La conjonction de ces deux itérations confère à cet algorithme une complexité quadratique ($O(n^2)$).

```

1 func allCombinations(
2   elements : array[Object],
3   result : array[array[Object]])
4 var i : int
5 var combination : array[Object]
6 begin
7   i ← 2
8   for i <= size(elements) do
9     combinations(elements, i, 0, combination, result)
10    i ← i+1
11  end for

```

Extrait de code 6.7 – Algorithme pour la génération d’arrangements de 2 à n éléments parmi n .

Ensuite, tous les arrangements sont parcourus, et, pour chacun, un concept de type `of4osm:IOSMTagCombinationConcept` est créé. L’algorithme 6.8 formalise ce processus. Afin de construire le label d’un `of4osm:IOSMTagCombinationConcept`, les concepts générés au cours de la phase de création de la hiérarchie clé-tag (section 6.3.3.1) sont utilisés. En effet, les concepts extraits des tags de type `of4osm:IOSMCategoryTag` existent déjà : il s’agit des concepts de type `of4osm:IOSMCategoryTagConcept`. Un concept `of4osm:IOSMTagCombinationConcept` représentant la conjonction de tags de type `of4osm:IOSMCategoryTag`, les concepts issus de ces derniers (`of4osm:IOSMCategoryTagConcept`) sont réexploités : la concaténation de leur label au format *Upper Camel Case* fait l’objet du label du `of4osm:IOSMTagCombinationConcept` représentant leur conjonction.

Par exemple, pour une instance décrite par l’intermédiaire de 3 tags A , B et C , des concepts de types `of4osm:IOSMCategoryTagKeyConcept` et `of4osm:IOSMCategoryTagConcept` correspondant aux tags A , B et C ont déjà été créés lors de la génération de la hiérarchie clé-tag. Dans cette nouvelle étape, ce sont les concepts correspondant aux arrangements (A,B) , (B,C) , (A,C) et (A,B,C) qui sont générés.

L’ordre des tags associés à un même objet géographique est indifférent ($(C,B,A) \equiv (C,A,B) \equiv (B,A,C) \equiv (B,C,A) \equiv (A,C,B) \equiv (A,B,C)$). Pour ne pas retenir de doublon, le tableaux contenant les parents de chaque concept aristotélien est trié par ordre alphanumérique¹⁸ (ligne 7 de l’algorithme 6.8). De cette manière, toutes les combinaisons des mêmes tags mais dont un ordre distinct ((C,B,A) , (C,A,B) , (B,A,C) , (B,C,A) , (A,C,B) , (A,B,C)) donnent lieu à la création d’un `of4osm:IOSMTagCombinationConcept` dont le label de préférence est ABC . Dans l’ontologie OF4OSM, les identifiants (IRI) des concepts aristotéliens sont construits sur la base de ce label : les doublons ne sont pas retenus.

Une fois le concept de type `of4osm:IOSMTagCombinationConcept` représentant la conjonction de n tags créé (lignes 10 et 11 de l’algorithme 6.8), et s’il n’existe pas déjà dans l’ontologie (ligne 12), des liens de subsomption sont établis entre lui et les concepts aristotéliens représentant la conjonction de $n-1$ tags (ligne 13).

18. Le choix du tri par ordre alphanumérique est arbitraire et n’est pas sans incidence sur la sémantique des concepts qui sont construits sur la base de cette approche. Les questionnements autour de ces enjeux linguistiques sont ici laissés de côté par souci de temps mais sont discutés dans les perspectives à nos travaux dans le dernier chapitre de ce manuscrit (cf. section 8.2).

Ainsi, le concept (A, B, C) , représentant la conjonction des concepts A , B et C , est fils des concepts (A, B) , (B, C) et (A, C) , lesquels représentent chacun la conjonction de 2 concepts. Enfin, le concept aristotélicien est ajouté à l'ontologie OF4OSM (ligne 14). Ainsi, chaque concept ne connaît que ses parents directs. Cependant, le caractère transitif de la relation de subsumption `rdfs:subClassOf`¹⁹ permettra, une fois l'ontologie générée, de reconstruire l'ensemble de la taxonomie grâce à des raisonneurs sémantiques tels que FaCT++ [Tsarkov and Horrocks, 2006], Pellet [Sirin et al., 2007] ou encore Hermit [Glimm et al., 2014].

```

1 func createAristoTaxoByElement(
2   onto : OF4OSMOntology,
3   categoryTagConcepts : array [IOSMCategoryTagConcept])
4 var combinations : array [array [IOSMCategoryTagConcept]]
5 var parents : array [IOSMCategoryTagConcept]
6 begin
7   sort(combinations)
8   allCombinations(categoryTags, combinations)
9   for each combination in combinations do
10    IOSMTagCombinationConcept aristoConcept
11    ← createOSMTagCombinationConcept(combination)
12    if !contains(onto, aristoConcept) then
13      addAllParents(aristoConcept, combination)
14      addConcept(onto, aristoConcept)
15    endif
16  end for
17 end

```

Extrait de code 6.8 – Algorithme pour la génération d'une hiérarchie de concepts issus de la conjonction de tags associés à un même objet géographique.

Sur une machine dotée d'un processeur de 2.2Ghz et de 8Go de RAM, le temps d'exécution de l'algorithme 6.8 sur les tags associés aux objets géographiques situés sur la ville de Grenoble est inférieur à 1 seconde. A l'issue de cette opération, 3 112 concepts aristotéliciens sont ajoutés à l'ontologie OF4OSM.

6.4 Conclusion

Dans ce chapitre, nous avons expliqué les méthodes et outils requis pour générer l'Ontologie extraite de la Folksonomie pour OSM. La section 6.2 a introduit le processus d'alimentation initiale de l'ontologie sur la base des tags répertoriés sur la page *Web Map Features* du wiki relatif au projet OpenStreetMap. La réalisation d'un client pour interroger le CMS sur lequel est basé ce wiki a été expliquée. Cet outil permet de générer automatiquement une première hiérarchie de concepts extraits

19. <https://www.w3.org/TR/rdf-schema/>

de tags largement utilisés et consensuels puisqu'ils ont préalablement été soumis à un processus de validation par la communauté OSM.

Par la suite, en section 6.3, nous avons vu comment l'ontologie OF4OSM est alimentée de façon complémentaire à partir de la base de données OSM. Les trois étapes de ce processus ont été détaillées. D'abord, nous avons décrit la phase d'extraction de l'information au moyen d'un client HTTP interrogeant un serveur conforme aux spécifications de l'API OSM (section 6.3.1). Puis, la méthode de caractérisation des tags OSM selon la typologie du méta-modèle de OF4OSM a été exposée (section 6.3.2). Enfin, la méthode de structuration des tags appuyée sur le modèle conceptuel du méta-modèle de OF4OSM a été présentée (section 6.3.3).

Le bilan quantitatif des concepts ajoutés à OF4OSM au cours de l'ensemble des processus décrits dans ce chapitre est illustré dans le tableau 6.2. L'alimentation initiale de l'ontologie via la page *Map Features* permet d'utiliser les tags les plus consensuels répertoriés sur le wiki OSM pour en extraire 26 concepts de type `of4osm:IOSMCategoryTagKeyConcept` et 898 concepts de type `of4osm:IOSMCategoryTagConcept`. Cette étape initiale permet également d'ajouter du lien entre certains de ces tags en les groupant sous 58 concepts de type `of4osm:IHighLevelConcept`. Ensuite, l'alimentation initiale ajoute, dans la phase de création de la hiérarchie de concepts issus des clés et des tags, 179 concepts de type `of4osm:IOSMCategoryTagKeyConcept` et 1 248 concepts de type `of4osm:IOSMCategoryTagConcept`. Cette étape permet de tenir compte des tags réellement utilisés par les contributeurs, au-delà des simples prescriptions du wiki OSM. Enfin, la deuxième phase d'alimentation complémentaire de l'ontologie OF4OSM via la base de données OSM, la classification aristotélicienne, crée 3 112 concepts issus de l'utilisation conjointe de tags OSM. Au total, après l'ensemble des processus décrits dans ce chapitre, l'ontologie OF4OSM contient 5 521 concepts organisés en une taxonomie dont la profondeur maximum est de 6 niveaux.

Dans le chapitre 7, nous verrons comment l'ontologie OF4OSM améliore significativement la qualité des données OSM au moment de l'acquisition mais également comment l'extraction de l'information s'en trouve facilitée.

TABLE 6.2 – Bilan quantitatif des concepts de l'ontologie OF4OSM après alimentations initiale et complémentaire

	Alimentation initiale (via page Web <i>Map Features</i>)		Alimentation complémentaire (via BD OSM)		Total
			Hierarchie clé-tag	Classification aristotélicienne	
of4osm:IHighLevelConcept	58	0	0	0	58
of4osm:IOSMCategoryTagKeyConcept	26	179	0	0	205
of4osm:IOSMCategoryTagConcept	898	1 248	0	0	2 146
of4osm:IOSMTagCombinationConcept	0	0	3 112	0	3 112
					5 521

Chapitre 7

Exploitation de l'ontologie OF4OSM

7.1 Introduction

Dans le chapitre 6, nous avons abordé la génération de l'ontologie OF4OSM. Dans un premier temps, le processus d'alimentation initiale de l'ontologie sur la base des informations disponibles sur la page *Map Features* du wiki dédié au projet OpenStreetMap a été expliqué (section 6.2). Sur cette page Web, les tags OSM et titres sous lesquels ils sont regroupés, sont parcourus par l'intermédiaire d'un client HTTP qui interroge l'API du CMS MediaWiki. Ensuite les concepts correspondant aux tags et titres sont créés puis ajoutés dans OF4OSM. Ces concepts étant extraits du guide de bonnes pratiques faisant référence pour l'utilisation des tags OSM, cette phase d'initialisation fournit une base conceptuelle solide à l'ontologie.

Dans un second temps, la méthode d'alimentation complémentaire de l'ontologie à partir des objets géographiques de la base de données OSM a été examinée. Celle-ci est divisée en 3 étapes. La première consiste en l'interrogation de la base de données par l'intermédiaire de l'API OSM via un client HTTP dédié. Ensuite, au cours de la seconde étape, les tags précédemment collectés sont caractérisés à l'aide de la typologie des tags OSM du méta-modèle de OF4OSM (cf. section 5.2.5). Enfin, lors de la troisième étape, des concepts sont créés et ajoutés à l'ontologie OF4OSM à partir des tags identifiés à l'étape précédente et selon le modèle conceptuel du méta-modèle de OF4OSM (cf. section 5.3).

Dans ce chapitre, nous montrons les avantages liés à l'exploitation de l'ontologie

OF4OSM qui résulte des différentes phases exposées au chapitre 6. D’une part, nous abordons les atouts de l’utilisation de OF4OSM pour l’amélioration de la qualité des données OSM au moment de l’acquisition (section 7.2). De l’autre, nous exposons les apports de cette nouvelle ontologie en ce qui concerne l’extraction de l’information recensée dans la base de données OSM (section 7.3).

7.2 Apports pour l’activité de tagging

Le contrôle de la qualité des données OSM au moment de l’acquisition, et, en particulier, l’amélioration de la pertinence des tags associés aux objets géographiques cartographiés, est crucial. Dans cette optique, l’exploitation de l’ontologie OF4OSM présente des avantages significatifs. Dans un premier temps, l’intérêt de cette ontologie pour l’optimisation des performances des systèmes de suggestions de tags similaires est abordé en section 7.2.1. Dans un second temps, les apports de l’ontologie pour enrichir les combinaisons de tags OSM associées aux objets de la base de données OSM sont expliqués en section 7.2.2. En section 7.2.3, les atouts de OF4OSM dans le processus d’identification des erreurs de tagging sont mis en avant. Dans la section 7.2.4, les avantages de l’exploitation de OF4OSM pour fournir un support actualisé pour la contribution à la folksonomie OSM sont montrés. Enfin, les bénéfices liés à une exploitation contextualisée de l’ontologie OF4OSM sont introduits en section 7.2.5.

7.2.1 Un étiquetage plus précis

Parmi les apports pour l’activité de tagging, l’ontologie OF4OSM lie les concepts issus des tags OSM avec une grande finesse, permettant aux éditeurs OSM de faire des suggestions de tags similaires plus précises au contributeur lorsqu’il ajoute un tag dans la base de données. En effet, au-delà de la relation de subsomption entre les concepts issus des clés des tags OSM et les concepts issus des tags qui partagent ces clés (cf. sections 5.3.1 et 6.3.3.1) déjà implémentée dans les travaux de [Auer et al., 2009] et [Codescu et al., 2011], l’ontologie OF4OSM intègre des concepts qui ne sont pas directement issus des tags OSM mais qui permettent de dresser des liens entre eux. Lors de l’initialisation de l’ontologie OF4OSM (cf. section 6.2), certains de ces concepts de type `of4osm:IHighLevelConcept` (cf. section 5.3.2) sont générés de façon automatique à partir des titres de niveau 3 de la page *Map features* du wiki OSM. Grâce à cette technique, en considérant que la similarité entre deux concepts de type `of4osm:IOSMCategoryTagConcept` est proportionnelle au nombre de parents

qu'ils ont en commun, on peut obtenir des mesures de similarité plus fines grâce à l'analyse de la structure de OF4OSM. L'utilisation de l'ontologie via un système de suggestion de tags similaires tel que celui utilisé par le greffon OSMantic¹ de l'éditeur OSM JOSM² permet de faire des propositions de tags plus pertinentes qu'en exploitant seulement la relation de subsomption entre clé et tag.

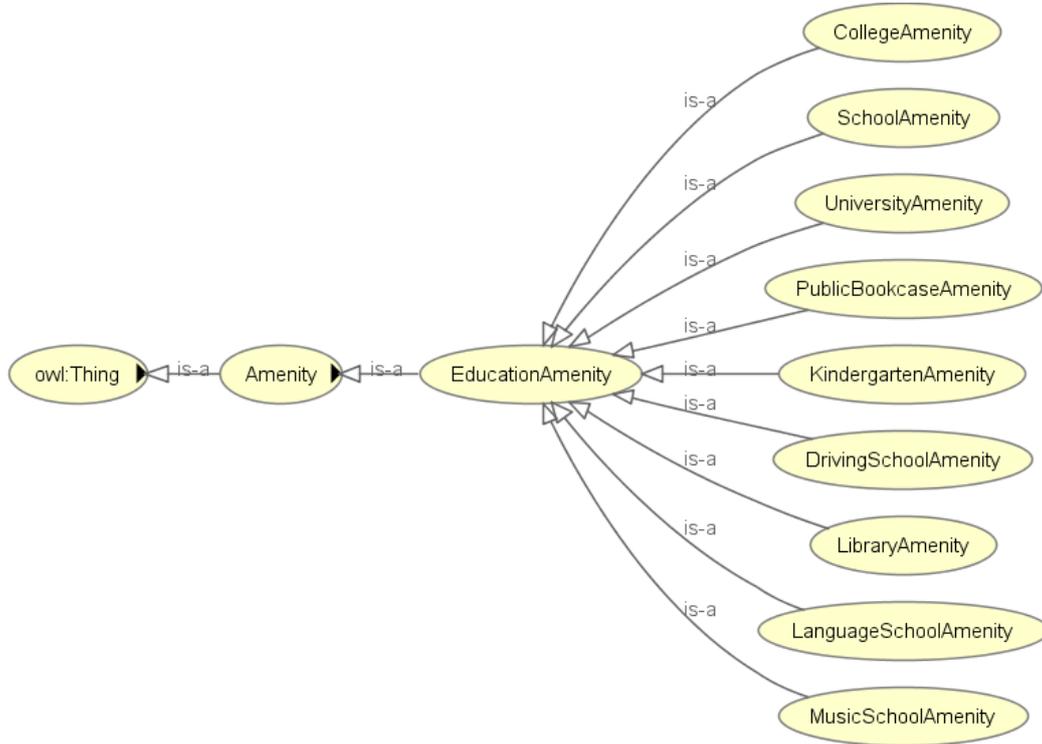


FIGURE 7.1 – Concept de haut niveau *Education*, concept parent des objets géographiques ayant trait à l'éducation (extrait de l'ontologie OF4OSM).

Par exemple, le concept de haut niveau (`of4osm:IHighLevelConcept`) *EducationAmenity*, illustré sur la figure 7.1 permet de lier plus étroitement les concepts fils du concept *Amenity* ayant trait à l'éducation. Ainsi, un système de suggestion de tags utilisant une mesure de similarité liée à la structure de l'ontologie OF4OSM affecte une plus grande similarité entre deux tags associés à des concepts fils de *EducationAmenity* (par exemple, les tags `amenity=school` et `amenity=kindergarten`, respectivement associés aux concepts *SchoolAmenity* et *KindergartenAmenity*) qu'entre deux tags associés à des concepts fils de *Amenity* qui ne partagent pas d'autre parent en commun (par exemple, `amenity=school` et `amenity=restaurant`, respectivement associés aux concepts *SchoolAmenity* et *RestaurantAmenity*). Par conséquent, les suggestions faites aux contributeurs OSM au moment

1. <http://wiki.openstreetmap.org/wiki/JOSM/Plugins/OSMantic>
 2. <https://josm.openstreetmap.de/>

de l'acquisition des données en base sont plus précises.

7.2.2 Un étiquetage plus complet

Tags complémentaires

Outre l'interconnexion plus étroite entre les concepts de l'ontologie OF4OSM, les concepts aristotéliens issus de la conjonction de plusieurs tags autour d'un même objet géographique (cf. sections 5.3.3 et 6.3.3.2) permettent de suggérer des tags à utiliser en combinaison avec d'autres tags. En effet, l'analyse des hiérarchies de concepts issues de la méthode de classification aristotélienne des instances présentes dans la base de données OSM suggère une similarité plus forte entre les tags associés à des concepts de type `of4osm:IOSMTagCombinationConcept`. Pour illustrer cet aspect, prenons les exemples des concepts *FortHistoricAttractionTourism* et *BookShopCafeAmenity* illustrés dans la figure 7.2.

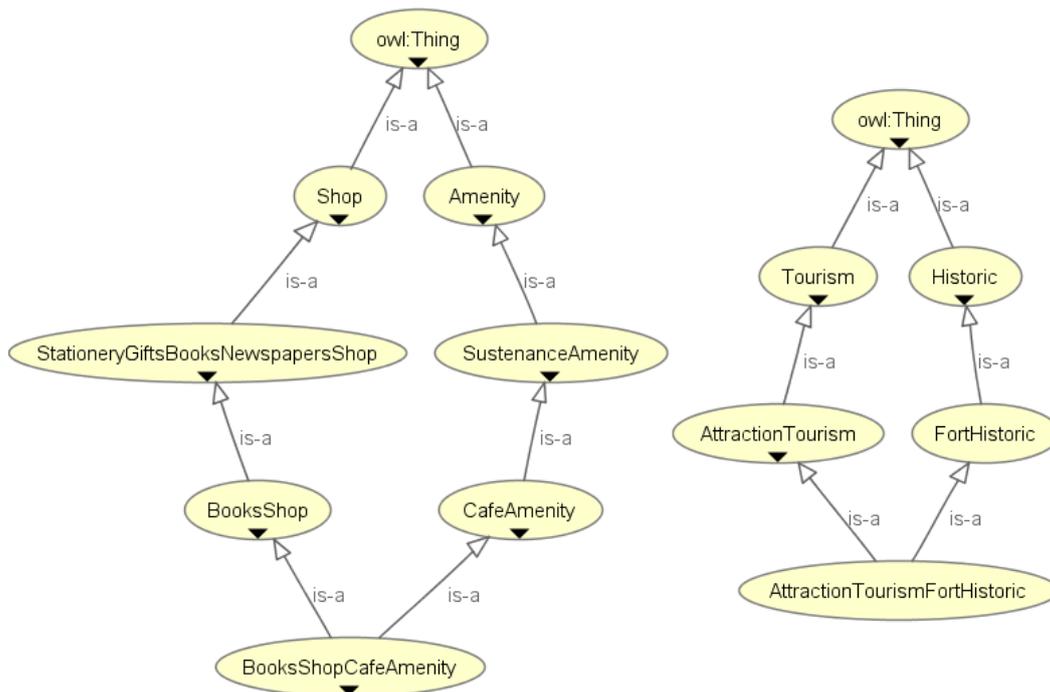


FIGURE 7.2 – Taxonomies issues de la classification aristotélienne d'un fort décrit comme une attraction touristique et d'une librairie dotée d'un café (extraits de l'ontologie OF4OSM).

Sur la partie gauche de la figure 7.2, il s'agit d'une hiérarchie de concepts construite à partir d'un objet géographique associé aux tags `historic=fort` et `tourism=attraction`. Sur la partie droite de la figure 7.2, il s'agit d'une hiérarchie

générée à partir d'un objet associé aux tags `shop=book` et `amenity=cafe`. Le premier objet est un fort qui est aujourd'hui un point d'intérêt pour les touristes tel que le fort de la Bastille à Grenoble. Le deuxième est une librairie disposant d'une partie café. Sur ces deux exemples, les concepts `of4osm:IOSMTagCombinationConcept` expriment la conjonction de deux fonctions. Dans le cas du fort, ces fonctions sont décalées dans le temps : une fonction passée (fortification militaire) et une fonction présente (attractivité touristique d'un territoire). Dans le cas de la librairie-café, les fonctions sont concomitantes.

Si l'on considère que la pertinence de la combinaison de deux tags est proportionnelle au nombre de concepts `of4osm:IOSMTagCombinationConcept` qu'ils partagent, alors l'analyse de la structure de OF4OSM permet des suggestions de combinaisons de tags plus fines. Grâce à cette analyse, lorsqu'un contributeur OSM cartographie un fort et lui associe le tag `historic=fort`, un système de suggestion de combinaisons de tags raisonnant sur la taxonomie de OF4OSM est en mesure de proposer le tag `tourism=attraction` au contributeur, et réciproquement. De la même manière, si un contributeur tague un objet `shop=book`, la même analyse permet de proposer la combinaison avec le tag `amenity=cafe`, et réciproquement. Cette approche améliore la complétude de la description des objets collectés dans la base de données OSM.

Il convient de noter que de telles suggestions peuvent être déroutantes. Dans l'optique de fournir une représentation la plus complète possible des aménités disponibles sur un territoire donné, les services, commerces et équipements qui sont faiblement représentés tels que les librairies-café ne doivent pas être exclus. Cependant, puisqu'une librairie est assez rarement un café, la suggestion de combinaison des tags `amenity=cafe` et `shop=book` doit en tenir compte. Pour répondre à ce besoin, chaque suggestion de combinaison peut être pondérée par le nombre de co-occurrences des tags utilisés conjointement. Cette information peut être récupérée par l'intermédiaire du service Web TagInfo³.

Blocs de tags

Par ailleurs, les concepts aristotéliens `of4osm:IOSMTagCombinationConcept` peuvent indiquer la conjonction de propriétés nécessaire à la représentation d'un type bien spécifique d'aménité. En effet, les tags ne se situent pas tous sur le même plan conceptuel. Par exemple, le concept *PitchLeisureClaySurfaceTennisSport* qui représente un terrain de tennis en terre battue illustré en figure 7.3 est le résultat de la conjonction des tags `sport=tennis`, `leisure=pitch` et `surface=clay`. Le premier

3. https://taginfo.openstreetmap.org/taginfo/apidoc#api_4_key_combinations

tag, `sport=tennis`, décrit un fait social : l'activité sportive du tennis. Le second tag, `leisure=pitch`, représente un type d'équipements de loisir : le terrain de sport. Le troisième tag, `surface=clay`, caractérise les propriétés physiques d'une surface : le revêtement en terre battue. Pour décrire un terrain de sport dont le revêtement est en terre battue et dédié à la pratique du tennis, l'ensemble de ces trois tags est requis. Pourtant, aucun des trois tags combinés, pris indépendamment, ne permet de décrire la nature d'un commerce, d'un service ou d'un équipement.

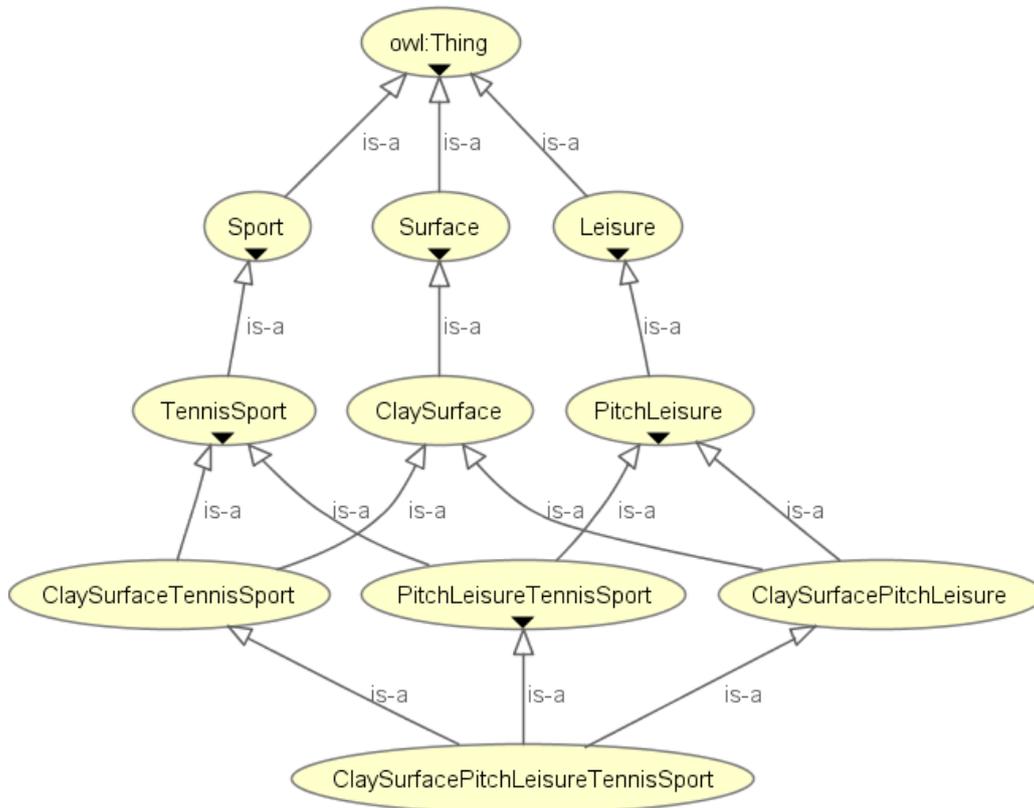


FIGURE 7.3 – Taxonomie issue de la classification aristotélicienne d'un terrain de tennis en terre battue (extrait de l'ontologie OF4OSM).

Les concepts parents de type `of4osm:IOSMCategoryTagConcept` d'un concept de type `of4osm:IOSMTagCombinationConcept` forment un ensemble correspondant à une combinaison de tags dans la base de données OSM. Cette représentation multi-dimensionnelle des objets géographiques par des concepts aristotéliciens permet de suggérer des combinaisons par blocs de plusieurs tags. Par exemple, les concepts *PitchLeisureClaySurfaceTennisSport*, *PitchLeisureAsphaltSurfaceBasketballSport* et *PitchLeisureEarthSurfaceSoccerSport*, présents dans l'ontologie OF4OSM, permettent de proposer à un contributeur qui tague un terrain `leisure=pitch` les combinaisons de tags `[surface=clay, sport=tennis]`, `[surface=asphalt, sport=basketball]` et

[`surface=earth`, `sport=soccer`]. Outre l'optimisation du temps de tagging, cette technique permet de réduire les incohérences de certaines combinaisons de tags en se saïssissant de la connaissance de la base OSM. En effet, d'après la façon dont les objets en base ont été taggés, le tag `sport=basketball` ne doit pas être combiné avec le tag `surface=earth`. Ce système de suggestion de combinaison multiples de tags améliore à la fois l'efficacité, la complétude et la cohérence de la description des entités géographiques de la base de données OSM. D'un autre côté, en proposant ces combinaisons, la description incomplète d'éléments spatialisés au moyen de tag dont l'emploi seul est insuffisant est déconseillée, voire évitée.

Les tags associés au terrain de tennis en terre battue sont répertoriés sur la page *Map features* du wiki d'OSM. On retrouve même les tags `leisure=pitch` et `surface=*` sur la page du wiki dédié au tag `sport=tennis`⁴. Cependant, les concepts aristotéliens de OF4OSM donnent une indication supplémentaire sur la vraisemblance de certaine combinaison. En effet, après analyse de la ville de Grenoble par notre approche, on ne trouve dans OF4OSM que des terrains de tennis dont le revêtement est de type terre battue (*PitchLeisureClaySurfaceTennisSport*), sable (*PitchLeisureSandSurfaceTennisSport*) ou asphalte (*PitchLeisureAsphaltSurfaceTennisSport*)⁵. Les valeurs `mud`, `ice` ou encore `salt` pour la clé `surface=*`⁶ ne seront donc jamais suggérées à un contributeur qui tague un terrain de tennis par un système de suggestions basé sur l'ontologie OF4OSM.

Combinaisons de tags extensive

Par ailleurs, les concepts `of4osm:IOSMTagCombinationConcept` expriment la conjonction des tags associés aux entités géographiques effectivement ajoutées à la base de données OSM par les contributeurs. Ils ne se cantonnent donc pas à la simple expression des combinaisons de tags documentés sur le wiki OSM. La figure 7.4 illustre deux taxonomies issues de la classification aristotélienne de tags parmi lesquels certains ne sont pas documentés sur le wiki. À droite sur la figure, il s'agit d'une hiérarchie de concepts extraite de la conjonction des tags `amenity=bar` et `music=reggae` associés à un bar diffusant de la musique de genre reggae. L'ambiance du bar semble être ici une caractéristique essentielle du commerce et donc légitime. Pour combler la lacune du wiki OSM, le contributeur qui identifie ce manque en parcourant l'arborescence de OF4OSM peut ajouter la clé `music=*` au guide de bonnes pratiques OSM et proposer la combinaison avec le tag `amenity=bar`.

4. <http://wiki.openstreetmap.org/wiki/Tag:sport>

5. [http://www.overpass-api.de/api/xapi?*\[bbox=5.677606,45.15414,5.753118,45.214077\]\[leisure=pitch\]\[surface=*\]](http://www.overpass-api.de/api/xapi?*[bbox=5.677606,45.15414,5.753118,45.214077][leisure=pitch][surface=*])

6. <http://wiki.openstreetmap.org/wiki/Key:surface>

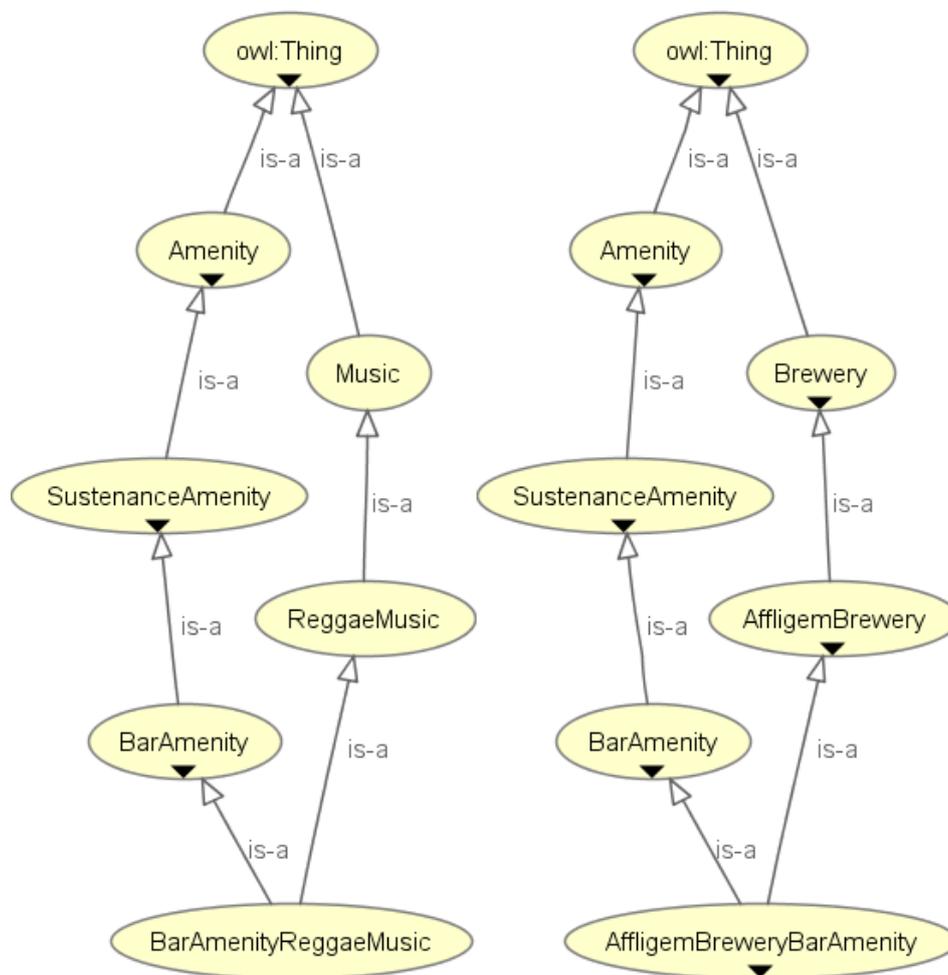


FIGURE 7.4 – Taxonomies issues de la classification aristotélicienne d’un bar diffusant de la musique Reggae et d’un bar proposant de la bière Affligem (extraits de l’ontologie OF4OSM).

À gauche sur la figure 7.4, il s’agit d’une hiérarchie de concepts issue de la classification aristotélicienne de la conjonction des tags `amenity=bar` et `brewery=affligem` associés à un bar proposant de la bière de marque Affligem à la pression. Le tag `amenity=bar` ainsi que la clé `brewery=*` sont effectivement répertoriés sur le wiki OSM. Cependant, les valeurs de la clé `brewery=*` ne sont pas listées sur le wiki OSM (probablement à cause de leur grand nombre ou parce que la publicité pour les marques est interdite). En extrayant les concepts `of4osm:IOSMTagCombination-Concept` à partir des objets en base de données, OF4OSM fait remonter les valeurs effectivement utilisées par les contributeurs. Ainsi, en utilisant OF4OSM, un système de suggestion de combinaison de tags peut proposer les tags dont les valeurs sont utilisées par les contributeurs OSM, sans qu’elles ne soient nécessairement do-

cumentées sur le wiki OSM.

7.2.3 Support pour l'identification d'erreurs de tagging

L'ontologie OF4OSM peut également être un support pour l'identification d'erreurs dans l'association des tags à un objet géographique. En effet, certains concepts `of4osm:IOSMTagCombinationConcept` expriment un mésusage de la folksonomie OSM. La figure 7.5 est un extrait de la taxonomie de OF4OSM du concept aristotélien *HackerspaceLeisureFablabWorkshop* et de ses parents. Ce concept représente un hackerspace-fablab, résultat de la conjonction des tags `leisure=hackerspace` et `workshop=fablab`. Cependant, l'utilisation conjointe de ces tags témoigne d'un recoupement sémantique dans la folksonomie OSM : un fablab est un type spécifique de hackerspace. Ici, OF4OSM peut être un outil adapté pour l'aide à la décision dans le but d'optimiser l'usage des tags OSM. En effet, à l'aide d'une interface graphique pour la visualisation d'ontologies telle que celle proposée par l'éditeur Protégé⁷, la revue des différents types d'entités géographiques caractérisés par la conjonction de plusieurs tags est aisée. En parcourant ces concepts aristotéliens, un contributeur peut facilement identifier les erreurs de tagging sans en avoir connaissance *a priori*.

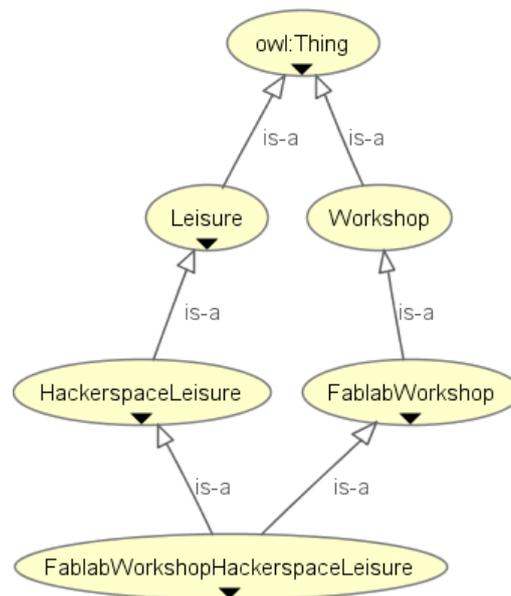


FIGURE 7.5 – Taxonomie issue de la classification aristotélienne d'un *fablab* (extrait de l'ontologie OF4OSM).

Dans l'optique d'éliminer le recoupement sémantique du hackerspace-fablab, le

7. <http://protege.stanford.edu/>

guide de bonnes pratiques sous forme de wiki peut être modifié par un contributeur afin d'indiquer aux autres utilisateurs le lien de spécialisation qui existe implicitement entre ces deux tags. Mieux, l'utilisateur peut créer un nouveau concept de haut niveau (`of4osm:IHighLevelConcept`) nommé *Third Place* pour désigner un "tiers-lieu", mot dérivé de l'acception selon laquelle le lieu de résidence est considéré comme le premier et le lieu de travail comme le second [Oldenburg, 1989]. Le concept *ThirdPlace* serait une super-classe des concepts *HackerspaceLeisure* et *FablabWorkshop* afin d'explicitier la caractéristique commune qui les lie. De cette façon, la navigation parmi les différents types de tags OSM serait facilitée (cf. section 7.3.1). D'autre part, le tag `workshop=fablab` pourrait être remplacé par le tag `hackerspace=fablab` afin d'explicitier le fait qu'un fablab est un type de hackerspace et qu'il ne peut exister d'objet géographique de type fablab qui ne soit pas d'abord tagué `leisure=hackerspace` avant d'être tagué `hackerspace=fablab`. L'exploitation de l'ontologie OF4OSM permet donc de décrire de façon plus exhaustive un objet de la base de données OSM en révélant les faiblesses de la folksonomie OSM et en offrant la possibilité de restructurer les tags au moyen de l'ontologie.

L'exemple illustré par la figure 7.6 est manifeste d'un autre cas d'utilisation inadéquate des tags OSM. Il s'agit cette fois d'une taxonomie issue de la classification aristotélicienne d'un bâtiment public qui abrite les bureaux d'un organisme gouvernemental. Elle est le fruit de l'analyse de la conjonction des trois tags `amenity=public_building`, `building=public` et `office=government`. Sur notre aire d'étude, la ville de Grenoble, la préfecture de l'Isère est l'entité géographique (la seule) à présenter l'ensemble de ces tags⁸. Plus que d'un simple chevauchement sémantique, il s'agit cette fois d'une redondance complète entre les tags `amenity=public_building` et `building=public`. Le guide de bonnes pratiques d'OSM proscriit l'usage du `amenity=public_building`. Grâce à l'exploration via Protégé des concepts aristotéliciens de l'ontologie OF4OSM, le contributeur OSM peut débusquer ce type de cas de redondance conceptuelle et le rectifier en base, en l'occurrence, en supprimant le tag `amenity=public_building` de la préfecture de l'Isère. OF4OSM peut donc servir de support pour identifier un mésusage des tags OSM.

7.2.4 Aide à la contribution actualisée

Plus de 90 000 personnes contribuent à la construction des 35 000 pages du wiki OSM⁹. Du côté de la base de données elle-même, entre 10 000 et 20 000 nouveaux

8. <https://www.openstreetmap.org/relation/53331>

9. <http://wiki.openstreetmap.org/wiki/Special:Statistics>, consulté le 15/11/2016.

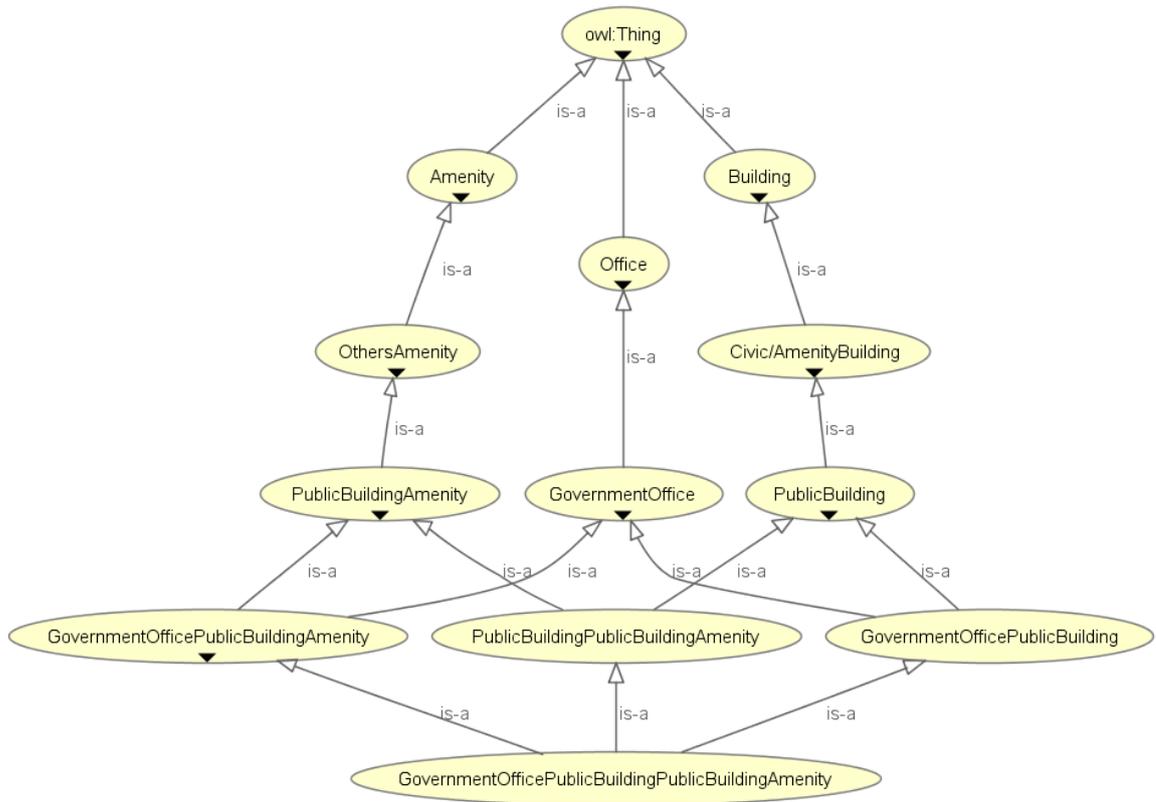


FIGURE 7.6 – Taxonomie issue de la classification aristotélicienne d’un bâtiment public qui abrite les bureaux d’un organisme gouvernemental (extrait de l’ontologie OF4OSM).

contributeurs rejoignent les rangs du projet chaque mois depuis début 2016¹⁰, et chacun d’entre eux modifie, en moyenne, 400 points par jour. L’édition d’un objet en base de données OSM concerne aussi bien sa géométrie, que son emprise spatiale ou encore les tags qui lui sont associés. Par conséquent, en ce qui concerne la folksonomie composée par l’ensemble des tags OSM, la fraîcheur de l’information est un enjeu de première importance.

De plus, les folksonomies ne regroupent qu’un ensemble de mots-clés, lesquels ne sont pas définis à l’avance (comme c’est le cas des nomenclatures). De ce fait, elles possèdent une expressivité formelle très faible comparativement aux autres types d’organisation de vocabulaires tels que les thesauri, les taxonomies ou encore les ontologies [Weller, 2007; Togia, 2005]. La contrepartie de cette expressivité formelle est la complexité de la structure. *A fortiori*, cette complexité est un obstacle majeur à l’évolutivité du vocabulaire qu’elle vise à structurer. Par conséquent, les folksonomies représentent un type de classification du contenu très faiblement structuré mais

10. <http://wiki.openstreetmap.org/wiki/Stats>, consulté le 15/11/2016.

également au potentiel évolutif très fort. Ce dernier est un atout majeur dans la cadre de projets participatifs tels qu'OpenStreetMap.

Pour tenir compte de l'évolutivité de la folksonomie OSM, l'ontologie OF4OSM est alimentée par deux processus : 1) le parcours et l'analyse de la page *Map Features* du wiki OSM expliqués en section 6.2 et 2) le parcours et l'analyse des objets géographiques de la base de données OSM abordés en section 6.3. Chacune de ces opérations est initiée par l'appel à des services Web : le premier processus a recours à l'API proposée par le CMS MediaWiki (cf. section 6.2.1), tandis que le second interroge un serveur qui implémente l'API Xapi (cf. section 6.3.1). L'invocation de ces services Web permet d'accéder aux données OSM au plus près de leur source et sans avoir à maintenir une copie de la base OSM. Sur la base de cette information actualisée, l'ontologie OF4OSM peut être fréquemment régénérée pour intégrer les données à jour. L'actualisation de l'ontologie ainsi facilitée permet d'accompagner et d'encourager l'effervescence qui anime l'activité de tagging dans OSM.

7.2.5 Aide à la contribution contextualisée

Les modèles et méthodes pour la génération de l'ontologie OF4OSM présentés dans les chapitres 3 et 6 sont génériques. Par conséquent, OF4OSM peut aussi bien être construite à partir de toute l'information disponible dans la base de données OSM, qu'à partir d'un extrait de la base de données correspondant à une boîte englobante donnée. Dans le cadre de nos expériences, nous avons choisi la ville de Grenoble comme aire d'étude. C'est à partir de cette zone que nous avons produit l'ontologie OF4OSM. La taxonomie illustrée en figure 7.7 est le résultat de la classification aristotélicienne de l'utilisation conjointe des tags `leisure=pitch`, `sport=boules` et `boules=lyonnaise` pour la description d'un terrain de boule lyonnaise représenté par le concept *BoulesSportLyonnaiseBoulesPitchLeisure*. La boule lyonnaise est une variante du jeu de boules communément appelée pétanque. Dérivée du jeu provençal, la pétanque est une spécificité régionale.

Dans un système de suggestion de tags (ou de combinaisons de tags), chaque suggestion peut être pondérée par le nombre d'occurrences de son objet (tag ou combinaison de tags) qui est un indicateur de sa popularité. Cependant, la distribution de ces occurrences n'est pas homogène sur le globe. En effet, les objets géographiques collectés dans la base de données OSM sont liés à des spécificités locales, comme dans le cas du terrain de boule lyonnaise représenté par la figure 7.7. Par exemple, la combinaison des tags `leisure=pitch` et `sport=boules` figurant

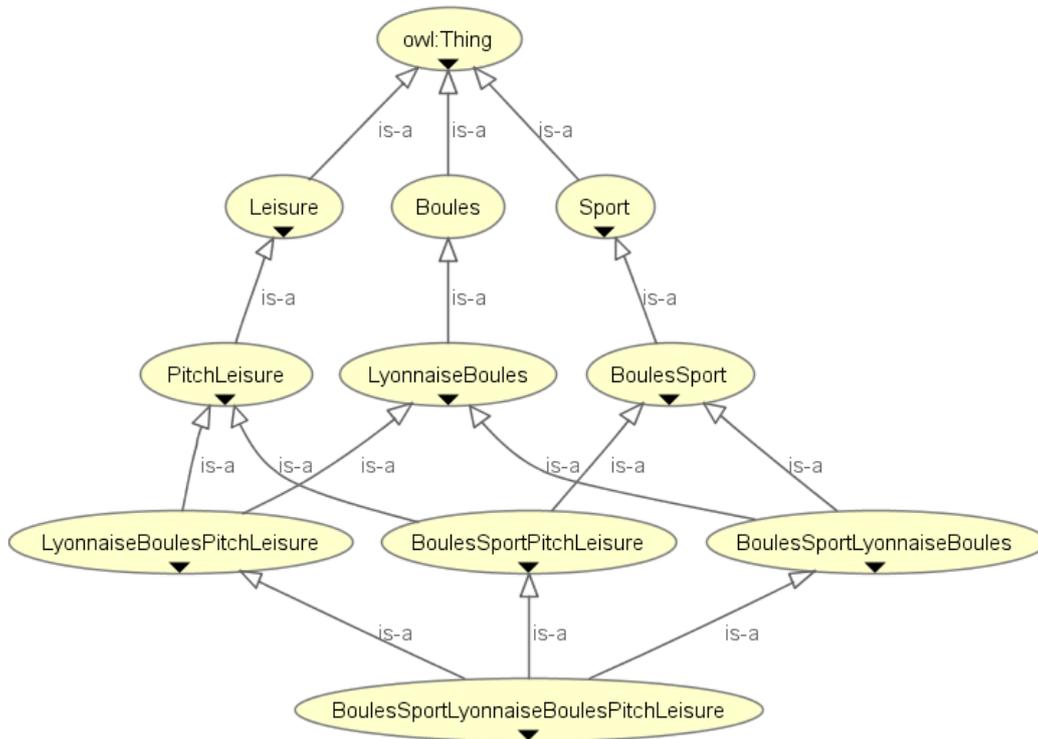


FIGURE 7.7 – Taxonomie issue de la classification aristotélicienne d’un terrain de jeu de boule lyonnaise (extrait de l’ontologie OF4OSM).

les terrains de pétanque n’est pas représentée dans la ville de Belfast en Irlande¹¹. À l’inverse, les terrains dédiés à la pratique des sports traditionnels gaéliques, représentés par la combinaison des tags `leisure=pitch` et `sport=gaelic_games` sont représentés à Belfast¹² mais pas à Marseille¹³. La suggestion du tag `sport=boules` en combinaison avec le tag `leisure=pitch` est pertinente si la contribution est située à proximité de la ville de Marseille, mais beaucoup moins à proximité de la ville de Belfast. De la même façon, la suggestion du tag `sport=gaelic_games` en combinaison avec le tag `leisure=pitch` est pertinente si la contribution est située à proximité de la ville de Belfast, mais beaucoup moins à proximité de la ville de Marseille.

Par conséquent, dans l’optique de proposer une aide efficace à l’activité de tagging, le nombre d’occurrences des tags (et de leur combinaison) ne doit pas être considéré globalement mais à une échelle plus fine. À cette fin, en procédant à une extraction partielle de la folksonomie OSM qui soit spécifique à la localisation de la contribution, la version locale de l’ontologie OF4OSM est une aide à la contri-

11. <http://overpass-turbo.eu/s/koo>, consulté le 15/11/2016.

12. <http://overpass-turbo.eu/s/kor>, consulté le 15/11/2016.

13. <http://overpass-turbo.eu/s/kov>, consulté le 15/11/2016.

bution contextualisée. En effet, dans ces versions locales de OF4OSM, les concepts ne sont extraits qu'à partir des tags ayant au moins une occurrence en base sur la BBOX considérée. Pour augmenter son efficacité, ce filtre spatial doit être couplé au nombre d'occurrences des tags liés aux concepts de type `of4osm:IOSMTagConcept` de l'ontologie. Le nombre d'occurrences d'un tag ou d'une combinaison de tags en base de données peut aisément être retrouvé via des services Web tels que l'API OverpassTurbo¹⁴.

7.3 Apports pour l'extraction de données

Au-delà des atouts qu'elle représente pour la phase d'acquisition des données, l'ontologie OF4OSM présente des bénéfices substantiels pour l'extraction de l'information contenue dans la base de données OSM. Dans un premier temps, l'intérêt de l'utilisation de OF4OSM pour la navigation entre les différents types d'objets géographiques cartographiés est abordé en section 7.3.1. Dans un second temps, les apports de l'ontologie pour rendre compte de l'aspect évolutif de la folksonomie OSM sont expliqués en section 7.3.2. Dans un troisième temps, les avantages de l'interopérabilité de OF4OSM pour obtenir l'information la plus exhaustive à propos des tags OSM sont exposés en section 7.3.3. Pour finir, les bénéfices d'une exploitation contextualisée de l'ontologie OF4OSM pour l'extraction d'information à partir d'une aire d'étude spécifique sont présentés en section 7.2.5.

7.3.1 Une typologie des objets géographiques plus fine

L'exploitation de l'ontologie OF4OSM permet aux utilisateurs de retrouver des objets géographiques dans la base de données OSM au moyen de concepts à la sémantique très fine. En effet, la méthode de construction de concepts aristotéliens, dont les principes sont exposés en section 5.3.3 et l'implémentation présentée en section 6.3.3.2, prend les objets de la base de données OSM comme vérité terrain pour construire une typologie de ces objets sur la base de la conjonction des tags qui leur sont associés. La mise en oeuvre de cette approche conduit à la production d'une taxonomie dont la profondeur exprime 1) le nombre de tags conjointement utilisés pour décrire un élément primitif (noeud/ligne/relation) spatialisé à laquelle s'ajoute 2) les deux niveaux de titre de la page *Map features* du wiki OSM (cf. section 6.4).

14. https://wiki.openstreetmap.org/wiki/Overpass_API

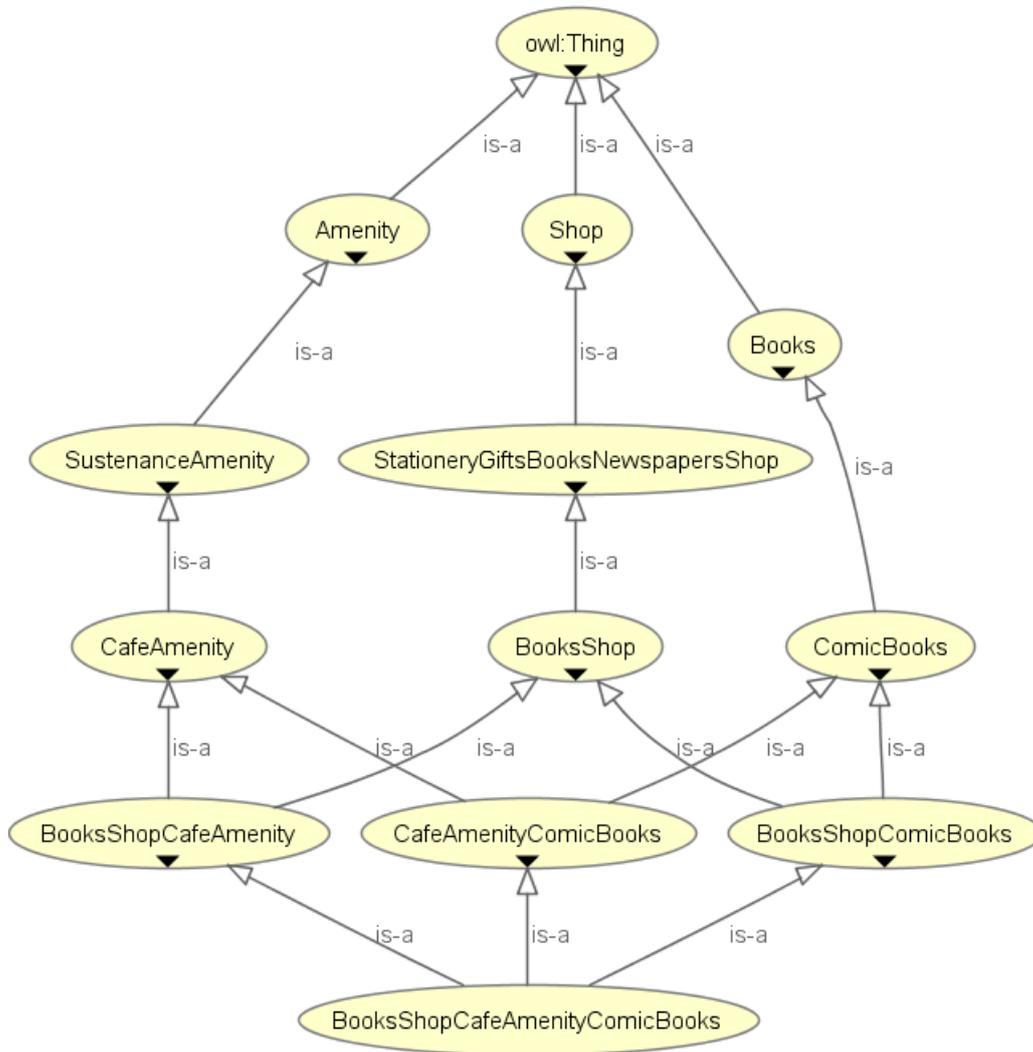


FIGURE 7.8 – Taxonomie issue de la classification aristotélicienne d’une librairie spécialisée dans les bandes dessinées et équipée d’un espace café (extrait de l’ontologie OF4OSM).

Par exemple, la figure 7.8 illustre la taxonomie issue de la classification aristotélicienne de l’utilisation conjointe des tags `shop=book`, `books=Comics` et `amenity=cafe`. Le concept *BooksShopCafeAmenityComicBooks* qui en résulte représente un magasin de livre spécialisé dans la vente de bandes dessinées et équipé d’une partie café. Ici, la hiérarchie extraite de la méthode de classification aristotélicienne des conjonctions de tags comporte 3 niveaux puisqu’elle convoque 3 tags OSM distincts. Les concepts *StationeryGiftsBooksNewspaperShop*¹⁵ et *SustenanceAmenity* sont eux issus du deuxième niveau de titre de la page *Map features*, respectivement

15. Ce concept est issu de l’étape d’initialisation de l’ontologie OF4OSM à partir de la page *Map Features*, il correspond au titre de niveau 3 *1.23.11 Stationery, gifts, books, newspaper*.

sous-niveaux des titres de premiers niveaux *Shop* et *Amenity* sur la même page. Au total, l'arborescence à 5 niveaux permet une navigation aisée au sein de la typologie des objets géographiques collectés dans la base de données OSM. En ce sens, OF4OSM représente un support à la navigation dans la folksonomie OSM.

7.3.2 Une ontologie révélatrice des pratiques émergentes

Les concepts de type `of4osm:IOSMTagCombinationConcept` sont issus d'une analyse *a posteriori* des tags collectés dans la base de données OSM. En ce sens, ils se distinguent des combinaisons décrites dans l'infobox des pages du wiki OSM sous l'intitulé "*useful combinations*". En effet, ces dernières visent à suggérer des tags dont l'usage est, *a priori*, fréquemment associé à l'usage d'autres tags. Par exemple, l'utilisation des tags `cuisine=*` et `diet=*` conjointement avec le tag `amenity=restaurant` est recommandé sur le wiki OSM, de telle sorte que la description d'un restaurant végétarien libanais se fait au moyen de l'utilisation des tags `amenity=restaurant`, `cuisine=libanese` et `diet=vegetarian`. Ces recommandations ont pour but de préciser la nature de l'objet géographique auquel la combinaison de tags est associée.

Les concepts aristotéliens peuvent, quant à eux, caractériser des pratiques émergentes dans l'aménagement des équipements et services sur le territoire. La figure 7.9 illustre deux exemples de cet aspect. La partie gauche représente la taxonomie issue de la classification aristotélienne des tags associés à un magasin de vêtements équipé d'un espace café. La partie droite représente la taxonomie issue de la classification aristotélienne des tags associés à un coiffeur spécialisé dans la culture punk. Etant donné leur nature extraordinaire, aucune de ces combinaisons ne fait l'objet d'une recommandation sur le wiki OSM.

Ces deux types de commerce ne reflètent qu'un nombre très restreint d'objets cartographiés dans la base de données OSM (deux instances de magasin de vêtement-café¹⁶ et une seule instance de coiffeur punk¹⁷ sur la ville de Grenoble). La méthode de classification aristotélienne permet de reconstituer une typologie des tags associés aux objets cartographiés dans la base de données OSM, *a posteriori*, c'est-à-dire à partir de la réalité terrain (telle qu'elle est cartographiée) et non de l'anticiper comme c'est le cas des recommandations de combinaisons indiquées sur le wiki OSM dans la rubrique *useful combinations*. À partir de l'ontologie OF4OSM, des combi-

16. [http://www.overpass-api.de/api/xapi?*\["bbox=5.677606,45.15414,5.753118,45.214077"\]&shop=clothes&amenity=cafe](http://www.overpass-api.de/api/xapi?*[), consulté le 15/11/2016.

17. [http://www.overpass-api.de/api/xapi?*\["bbox=5.677606,45.15414,5.753118,45.214077"\]&shop=hairstylist&culture=punk](http://www.overpass-api.de/api/xapi?*[), consulté le 15/11/2016.

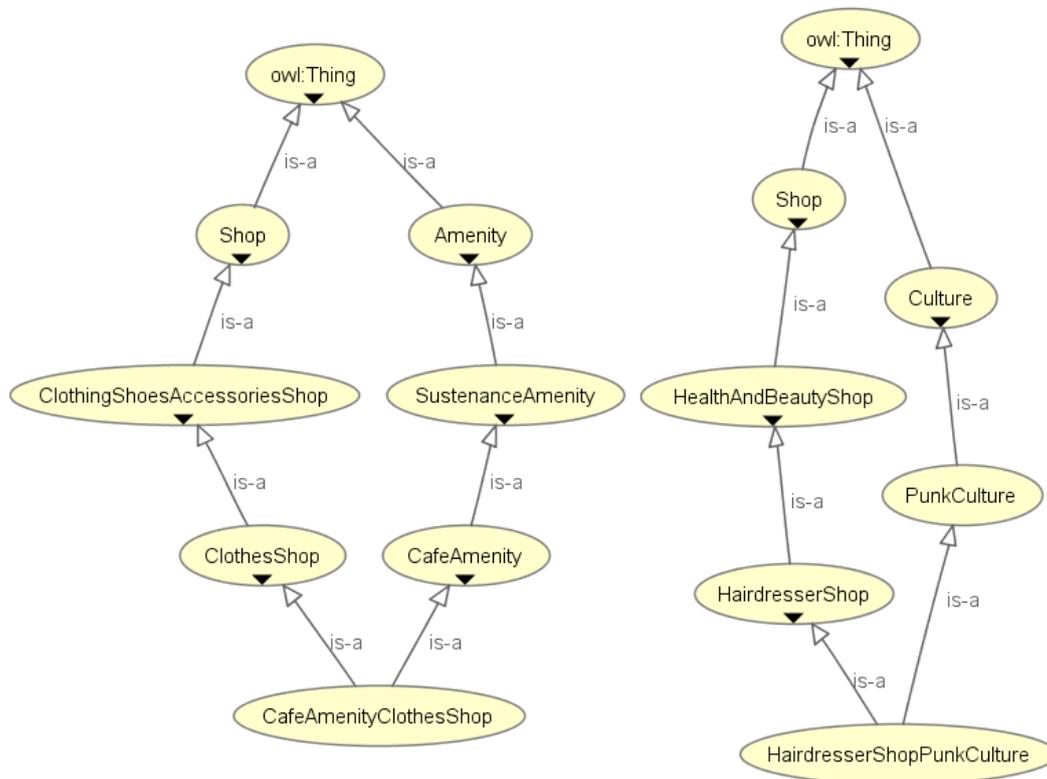


FIGURE 7.9 – Taxonomies issues de la classification aristotélicienne d’une librairie équipée d’un espace café et d’un coiffeur spécialisé dans la culture punk (extraits de l’ontologie OF4OSM).

naisons aussi insolites que celles évoquées plus haut peuvent être exploitées, dès lors qu’il en existe au moins une instance sur la zone géographique étudiée.

Du fait même de leur petit nombre, l’utilisation de ces combinaisons est limitée. Dans le cadre de la navigation dans l’arborescence de l’ontologie, il nous paraît judicieux d’indiquer le nombre d’occurrences des concepts `of4osm:IOSMTag-CombinationConcept`. Le nombre d’occurrences peut facilement être retrouvé en interrogeant l’API Xapi. Par exemple, les requêtes HTTP suivantes renvoient respectivement les objets géographiques correspondants aux concepts *ClothesShopCafeAmenity* et *HairdresserShopPunkCulture* situés sur la ville de Grenoble (BBOX : 5.677606, 45.15414, 5.753118, 45.214077) :

```

1 GET http://www.overpass-api.de/api/xapi?
2   * [ bbox=5.677606,45.15414,5.753118,45.214077 ]
3   [ shop=hairdresser ] [ culture=punk ]
4
5 GET http://www.overpass-api.de/api/xapi?
6   * [ bbox=5.677606,45.15414,5.753118,45.214077 ]
7   [ shop=clothes ] [ amenity=cafe ]

```

Malgré un nombre d'instances faible, l'ontologie OF4OSM relaie l'existence d'initiatives marginales sur le territoire afin de ne pas étouffer le développement de nouvelles expériences urbanistiques. En tenant compte de toute contribution, de signaux faibles mais pertinents, notre approche est en adéquation avec la philosophie participative du projet OSM. L'ontologie OF4OSM est le résultat d'une analyse de la folksonomie OSM telle qu'elle est - en récupérant directement l'état de la base à un moment donné - et non telle qu'elle devrait être selon les principes établis dans le wiki OSM. Cet aspect participatif génère une représentation de l'espace qui exprime le point de vue des experts terrain que sont les contributeurs OSM. Ainsi, la flexibilité inhérente à la folksonomie OSM se répercute sur l'ontologie OF4OSM.

7.3.3 Une interopérabilité accrue

La majeure partie des concepts de l'ontologie OF4OSM a été créée à partir de la folksonomie OSM. En effet, chaque concept qui instancie une des classes `of4osm:IOSMCategoryTagKeyConcept`, `of4osm:IOSMTagConcept` ou encore `of4osm:IOSMTagCombinationConcept` du modèle conceptuel de OF4OSM est créée à partir d'un ou plusieurs tags. Le lien entre chacune de ces instances et le ou les tag(s) sur le(s)quel(s) elle a été conçue est conservé via la propriété `skos:related` qui a pour rang un `of4osm:IOSMTag`. Ce dernier est identifié par une IRI correspondant à l'URL de description du tag sur le wiki OSM.

La persistance de l'information sur l'origine d'un concept issu de la folksonomie OSM garantit l'interopérabilité de l'ontologie OF4OSM avec d'autres bases de connaissances qui identifient les tags OSM de la même manière. C'est par exemple le cas de l'ontologie LGD qui référence aussi bien les tags OSM que les entités géographiques décrites dans la base de données OSM [Auer et al., 2009]. Les IRI du wiki OSM sont également utilisées pour identifier les tags au sein du *OSM Semantic Network*. Il est ainsi possible de retrouver les mesures de similarité entre tags produites par les auteurs de [Ballatore et al., 2013]. L'interopérabilité de l'ontologie OF4OSM offre la possibilité aux contributeurs OSM de profiter du maximum d'informations disponibles sur les tags OSM et les entités géographiques auxquelles ils sont associés.

7.3.4 Une exploration contextualisée

Dans le cadre d'une exploration de la base de données OSM focalisée sur une zone donnée plutôt que considérée dans son ensemble, l'extraction d'une version spatialement contextualisée de l'ontologie OF4OSM peut s'avérer profitable. En effet,

si un contributeur souhaite retrouver de l'information sur une aire d'étude donnée, il est possible, selon la méthode exposée au chapitre 7, de générer des concepts à partir de tags associés aux objets situés sur une BBOX donnée. L'ontologie OF4OSM ainsi créée comporte uniquement des concepts qui sont liés à des tags qui existent sur la zone d'étude. Par conséquent, le nombre de concepts contenus dans l'ontologie est moindre donc son exploration s'en trouve facilitée. Ce filtre spatial permet d'obtenir une base de connaissances OSM recentrée sur une zone géographique spécifique. La figure 7.10 est un extrait de l'ontologie OF4OSM produite à partir d'une extraction des données OSM sur la ville de Grenoble. Elle révèle des exemples symptomatiques de concepts spécifiques à une région que l'on trouve dans une version contextualisée de l'ontologie OF4OSM.

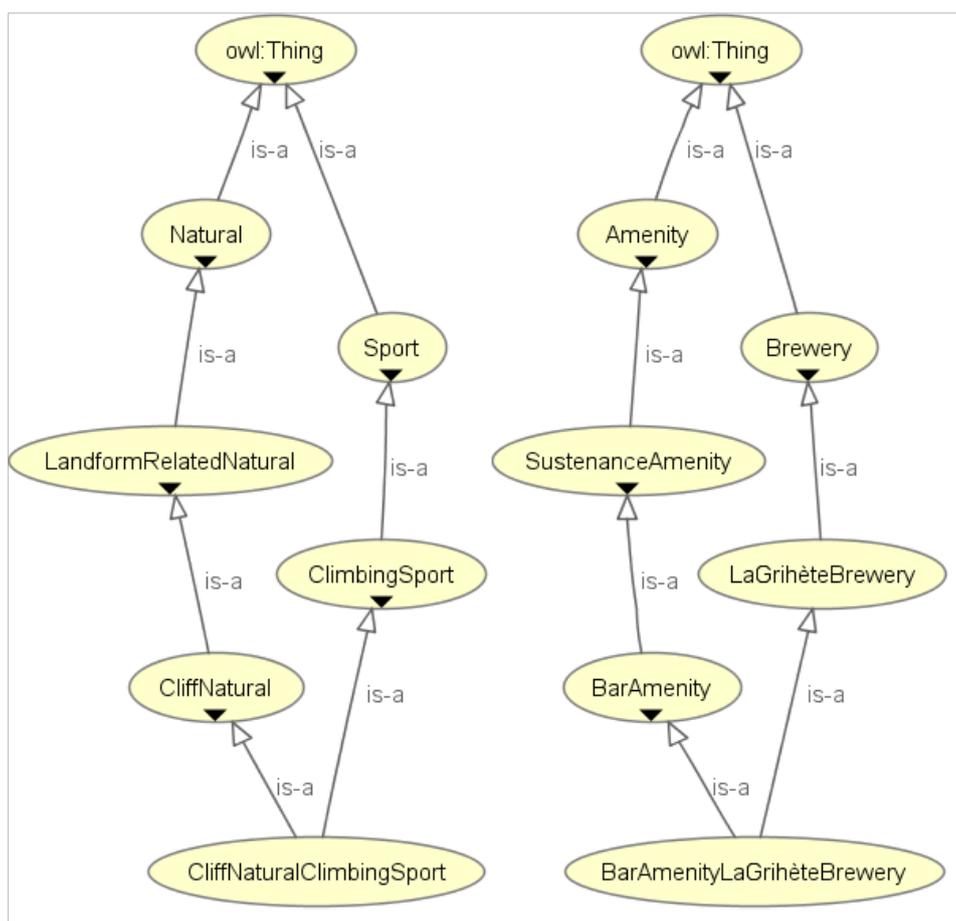


FIGURE 7.10 – Taxonomies issues de la classification aristotélicienne d'une voie d'escalade et d'un bar proposant de la bière de marque La Grihète (extraits de l'ontologie OF4OSM).

Le concept de type `of4osm:IOSMTagCombinationConcept` *BarAmenityLaGrihèteBrewery* représenté dans la partie droite de la figure 7.10 provient de la classifica-

tion aristotélécienne des tags `amenity=bar` et `brewery=LaGrihete`, utilisés conjointement pour décrire un bar proposant de la bière pression produite par la brasserie La Grihète. Cette brasserie d'une taille modeste est située dans le département de la Drôme, département voisin de l'Isère. Ceci explique que certains bars de Grenoble, chef-lieu de l'Isère, servent de la bière La Grihète à la pression. Le concept de type `of4osm:IOSMTagCombinationConcept` *CliffNaturalClimbingSport* représenté dans la partie gauche de la figure 7.10 provient de la classification aristotélécienne des tags `natural=cliff` et `sport=climbing`, utilisés conjointement pour décrire une falaise reconnue pour la pratique de l'escalade. La distribution de ces entités géographiques dans la base de données OSM n'est pas homogène sur le territoire. La version de l'ontologie OF4OSM spécifique à la ville de Paris ne contient ni le concept *BarAmenityLaGriheteBrewery* ni le concept *CliffNaturalClimbingSport* puisqu'aucun objet n'y est associé aux combinaisons de tags `[natural=cliff,sport=climbing]`¹⁸ ou `[amenity=bar,brewery=LaGrihete]`¹⁹. L'exploitation d'une version spatialement contextualisée de l'ontologie OF4OSM permet d'aider un utilisateur à explorer les différents types d'objets situés dans la base OSM sur une région particulière sans être noyé dans le bruit de concepts qui ne sont pas représentés sur cette même région.

7.4 Conclusion

Au cours de ce chapitre, nous avons passé en revue les multiples atouts de l'exploitation de l'ontologie OF4OSM. Ceux-ci ont trait à deux phases dans la chaîne de traitement de l'information dans le projet OpenStreetMap : l'acquisition et l'extraction. En ce qui concerne l'acquisition de la données, OF4OSM améliore cinq aspects. D'abord la précision des tags utilisés grâce à l'expressivité formelle de son méta-modèle (section 7.2.1). Ensuite, l'ontologie est en elle-même une base de connaissances sur l'utilisation conjointe des tags OSM grâce à sa méthode de classification aristotélécienne (section 7.2.2). De plus, OF4OSM est une aide précieuse dans le processus d'identification des mésusages de la folksonomie OSM car elle permet d'avoir une vision globale et facile d'accès de la façon dont sont décrits les objets géographiques en base, permettant par là-même de débusquer certaines défaillance (section 7.2.3). Par ailleurs, grâce à son processus de récupération de la donnée par l'intermédiaire de services Web, l'ontologie OF4OSM est un support à l'activité de tagging qui embarque les recommandations du guide de bonnes pratiques actualisées (section 7.2.4). Enfin, la méthode de génération de OF4OSM permet d'obtenir des versions de l'ontologie qui rendent compte des spécificités locales et fournit ainsi une

18. <http://overpass-turbo.eu/s/kB0>, consulté le 15/11/2016.

19. <http://overpass-turbo.eu/s/kom>, consulté le 15/11/2016.

aide à la contribution contextualisée (section 7.2.5).

À propos de la restitution de l'information à partir de la base de données OSM, l'ontologie OF4OSM apporte quatre améliorations significatives. La première concerne la navigation au sein de la typologie des objets géographiques collectés dans la base de données OSM (section 7.3.1). En effet, le méta-modèle de OF4OSM permet de générer une taxonomie d'une grande profondeur, offrant à l'utilisateur une vision structurée de la folksonomie OSM. Ensuite, en extrayant les tags directement depuis la base OSM, et non seulement à partir du guide de bonnes pratiques, l'ontologie OF4OSM rend compte des évolutions du territoire qui ne sont forcément prévisibles et donc pas (encore) documentées dans les recommandations du wiki OSM (section 7.3.2). D'autre part, grâce à son méta-modèle conforme aux standards du sémantique Web, mais également grâce à une identification compatible avec les différentes bases de connaissances sur les tags, l'exploitation de l'ontologie OF4OSM permet au contributeur OSM de profiter de multiples sources d'information lors de sa navigation dans la base de données OSM (section 7.3.3). Enfin, l'exploration de l'information OSM sur une zone d'étude précise est facilitée par la possibilité de générer des versions de l'ontologie OF4OSM contextualisée qui soulignent les spécificités régionales (section 7.3.4).

Chapitre 8

Bilan et perspectives

Au travers de nos travaux de thèse, nous avons proposé des réponses à la problématique suivante : comment augmenter la puissance sémantique de la folksonomie OSM tout en préservant la flexibilité inhérente à l'activité de tagging ? Pour cela, nous avons dressé, dans une première partie, un état de l'art des efforts pour la structuration des tags OSM. Dans une seconde partie, nous avons introduit un méta-modèle et des méthodes pour la structuration des tags en une base de connaissances OSM sous la forme d'une ontologie : l'ontologie OF4OSM.

Dans ce chapitre, nous faisons, dans un premier temps, le bilan de nos contributions et nous mettons en lumière les réponses apportées pour dépasser les limites des modèles de structuration de la folksonomie OSM revus dans l'état de l'art (section 8.1). Dans un second temps, nous abordons des perspectives de travaux futurs qui s'inscrivent dans le prolongement de nos recherches (section 8.2).

8.1 Bilan des réalisations

Dans ce manuscrit, nous avons proposé **un méta-modèle pour la structuration de la folksonomie OSM**. Ce méta-modèle permet d'extraire **une base des connaissances OSM sous la forme d'une ontologie que nous appelons OF4OSM**. Le méta-modèle de OF4OSM est composé de deux parties. La première supporte la caractérisation des tags en proposant **une représentation spécifique des tags OSM** et de leur composantes - la clé et la valeur (cf. section 5.2). Cette représentation étend une ontologie de référence pour l'activité de tagging (l'ontologie MUTO) afin de proposer une représentation des tags OSM qui soit intégrée

dans l'écosystème des vocabulaires pour la représentation des tags. Sur la base de cette représentation, et en fonction du type de leurs composantes, les tags sont classifiés selon **une typologie des tags OSM** que nous avons élaborée. Le deuxième partie du méta-modèle est le schéma conceptuel qui structure l'ontologie OF4OSM (cf. section 5.3). Ce schéma organise dans l'ontologie OF4OSM les concepts extraits des tags OSM identifiés au moyen de notre typologie. Ce schéma supporte trois méthodes de structuration de la connaissance OSM à partir des tags. La première consiste en une **hiérarchisation de concepts extraits des clés des tags et de concepts extraits des tags eux-mêmes**. La deuxième a trait à la **factorisation de caractéristiques partagées par certains tags en des concepts de haut niveau**. Enfin, le troisième volet repose sur la **création de concepts aristotéliens correspondants à la conjonction de différents tags associés à un même objet géographique**.

Par la suite, nous avons implémenté ces modèles et méthodes dans **un outil pour la génération automatique de l'ontologie OF4OSM**. Ce processus est composé de deux phases. La première est la phase d'**alimentation initiale de l'ontologie sur la base des tags répertoriés dans le guide de bonnes pratiques du projet OpenStreetMap** (cf. section 6.2). Ce guide, sous la forme d'une page de wiki, est exploré par un client Web dédié que nous avons développé. Le client interroge le CMS sur lequel est basé le wiki et récupère les tags OSM qui y sont répertoriés. À l'issue de ce processus, l'ontologie OF4OSM contient une hiérarchie de concepts extraits de tags OSM largement utilisés et consensuels puisqu'ils ont préalablement été soumis à un processus de validation par la communauté OSM au moyen du wiki. La seconde phase correspond à l'**alimentation complémentaire de l'ontologie OF4OSM à partir de la base de données OSM** (cf. section 6.3). Les trois étapes de ce processus ont été détaillées : 1) extraction de l'information au moyen d'un client HTTP interrogeant un serveur conforme aux spécifications de l'API OSM, 2) caractérisation des tags OSM selon la typologie du méta-modèle de OF4OSM, 3) structuration des tags appuyée sur le modèle conceptuel du méta-modèle de OF4OSM. À la fin de ces deux phases, l'ontologie OF4OSM contient plus de 5000 concepts organisés en une taxonomie dont la profondeur maximale est de 6 niveaux.

Les principaux apports de l'ontologie OF4OSM sont les suivants :

Couverture des tags OSM

À l'instar de l'approche GWAP, la génération de l'ontologie OF4OSM supportée par notre méta-modèle peut potentiellement extraire l'ensemble des tags de la base de données OSM. En effet, bien que, dans le cadre de notre étude,

la zone géographique considérée correspond à la ville de Grenoble, **notre approche est générique : elle est applicable à n'importe quelle aire d'étude, voire à l'ensemble de la base de données OSM**. Par conséquent, l'ontologie OF4OSM est capable de couvrir l'ensemble de la folksonomie OSM.

Expressivité formelle

Ensuite, OF4OSM bénéficie d'une expressivité formelle riche sous la forme des trois volets qui composent son schéma conceptuel : 1) la hiérarchie clé-tag, 2) les concepts de haut niveau et 3) la taxonomie aristotélicienne. Outre ces **quatre classes et les relations de subsomption (rdfs:subClass)** qui les lient, les **restrictions sur la qualité de ces relations de subsomption** sont également décrites dans le méta-modèle de OF4OSM. Cet ensemble de règles confère à l'ontologie OF4OSM les moyens d'une représentation des tags à l'expressivité significative.

Interopérabilité

Par ailleurs, la conformité du méta-modèle de OF4OSM par rapport aux standards favorise l'interopérabilité des systèmes et donc la réutilisation des données qui en sont issues [Salzano, 2013]. En effet, **notre ontologie est une extension de l'ontologie MUTO, elle-même proposant une représentation fédérée de différents vocabulaires pour la représentation de l'activité de tagging**. Cette conformité permet à l'ontologie OF4OSM d'être interopérable avec l'essentiel des efforts pour la structuration des tags OSM revus dans notre état de l'art. Cette compatibilité offre la possibilité de pouvoir comparer la représentation des connaissances OSM présentées dans notre ontologie à d'autres bases.

Implication des utilisateurs

Pour finir, l'ontologie OF4OSM est en phase avec l'aspect participatif du projet OpenStreetMap en cela qu'il ne tient pas simplement compte des tags qui font autorité sur le wiki OSM mais **considère les tags effectivement utilisés par les contributeurs OSM pour décrire les objets géographiques collectés dans la base de données OSM**. La prise en compte de tous les tags et la possibilité de générer automatiquement la base de connaissances qui en est issue, confère à OF4OSM une flexibilité prolongeant celle de la folksonomie OSM. De plus, grâce à sa méthode de génération automatique basée sur des services Web qui fournissent les données OSM à jour, **l'ontologie OF4OSM peut aisément être synchronisée avec le guide de bonnes pratiques d'OSM, ainsi qu'avec sa base de données**. L'ontologie OF4OSM peut donc accompagner l'évolution du projet OpenStreetMap.

8.2 Perspectives

Dans le prolongement des travaux de thèse présentés dans ce manuscrit, plusieurs perspectives de recherche sont envisagées.

Greffon éditeur OSM

La première d'entre elles réside dans le développement d'un greffon pour éditeur de données OSM. En effet, ce plugin doit d'abord permettre d'**explorer les données via le graphe dans lesquels s'organisent les concepts de l'ontologie OF4OSM**, ceci afin de tirer parti des avantages de OF4OSM pour l'extraction de données (cf. section 7.3). De plus, afin d'exploiter l'ontologie OF4OSM comme appui pour le contrôle de l'information au moment de l'acquisition, **un système de suggestion de tags similaires basé sur la structure de OF4OSM** peut être réalisé. Dans ce plugin, les différentes mesures de similarité entre les tags OSM abordées en section 7.2 seront implémentées. Lors de l'activité de tagging, ces mesures permettront de suggérer aux contributeurs des tags plus précis ou des combinaisons de tags afin d'améliorer la précision et la complétude de la description des objets géographiques de la base OSM. Parmi les travaux autour de la gestion et de la qualité des données issues du projet OSM abordés dans [Arsanjani et al., 2015], les auteurs de [Vandecasteele and Devillers, 2015] exposent une approche qui va dans ce sens, et qu'ils ont implémentée dans un greffon, OSMantic¹, pour l'éditeur OSM JOSM². Ce plugin pourrait être une référence pour réaliser le couplage entre l'ontologie OF4OSM et la base de données OSM.

Tests utilisateurs

Grâce à l'implémentation de cet outil, des tests utilisateurs pour l'évaluation de l'efficacité de l'aide à la contribution que représente l'ontologie OF4OSM devront être effectués. Les avantages de l'exploitation de l'ontologie OF4OSM portent à la fois sur la phase d'acquisition des données et sur la phase d'extraction de ces mêmes données. Par conséquent, les tests auront pour objet l'évaluation de l'apport de l'utilisation de OF4OSM, relativement à ces deux aspects. En ce qui concerne l'acquisition des données, **l'analyse des résultats de l'activité de tagging par un échantillon représentatif de contributeurs OSM assistés par le greffon basé sur OF4OSM** révélera les atouts et limites de notre approche. Du côté de l'extraction de l'information OSM, **un questionnaire adressé aux utilisateurs de la base de données OSM** nous permettra d'obtenir des retours sur l'intérêt de l'utilisation

1. <http://wiki.openstreetmap.org/wiki/JOSM/Plugins/OSMantic>

2. <https://josm.openstreetmap.de/>

de l'ontologie OF4OSM pour l'exploration de l'information OSM. L'ensemble de cette démarche mettra en évidence les forces et faiblesses de notre approche dans le but d'améliorations futures.

Méthodes d'évaluation de la qualité des tags

Par ailleurs, des méthodes d'évaluation de la qualité des tags OSM en amont de leur intégration dans l'ontologie OF4OSM sont à élaborer. Il est patent que tous les objets géographiques de la base de données OSM ne sont pas tous taggés avec la même acuité (précision, complétude, cohérence, etc.). Pourtant, nous avons proposé des méthodes et modèles qui exploitent indifféremment toutes les entités géographiques. Une **typologie des problèmes de qualité liés à l'utilisation des tags OSM** doit donc être développée pour identifier précisément les insuffisances de la folksonomie OSM. Ensuite, des **méthodes d'intégration des tags OSM dans l'ontologie OF4OSM en fonction de leur qualité** permettront d'obtenir une ontologie de qualité supérieure. Dans cette optique, le système d'identification d'erreurs dans l'utilisation de tags proposé par le projet OSMose³ peut être un point de départ. De plus, les nombreux travaux qui se sont intéressés à l'élaboration de méthodes d'évaluation de la qualité des contributions dans le domaine de la VGI [Kekler et al., 2013; Bégin et al., 2013; Rehrl et al., 2013; Bishr, 2011; Exel and Dias, 2010] peuvent apporter un éclairage complémentaire sur ces problématiques.

Intégration des spécificités vernaculaires

D'autre part, en l'état, les modèles et méthodes que nous avons proposés pour la génération automatique de l'ontologie OF4OSM tiennent compte uniquement de la langue anglaise. En effet, le wiki OSM préconise l'utilisation de l'anglais dans la construction des tags, c'est pourquoi nous avons utilisé cette langue pour le nommage des concepts qui composent notre ontologie. Néanmoins, la génération de OF4OSM à partir des tags associés aux objets géographiques situés sur la ville de Grenoble a mis en évidence l'existence de particularismes régionaux, et ce à différentes échelles (cf. sections 7.2.5 et 7.3.4). Dans l'optique d'offrir la possibilité aux contributeurs OSM de prendre en compte les spécificités vernaculaires de leur environnement, des **modèles et méthodes pour la génération multilingue de l'ontologie OF4OSM** devront être élaborés. Ici, les techniques d'analyse linguistique pour l'enrichissement d'ontologies topographiques élaborées par les auteurs de [Sébastien et al., 2011] seront une source d'inspiration.

Système de nommage multilingue

De plus, le méta-modèle de OF4OSM définit des labels qui servent à conserver la traduction des concepts initialement nommés en anglais. Cependant, dans le

3. <http://wiki.openstreetmap.org/wiki/Osmose>

cas où certains concepts (et donc certains tags) existent dans une région et pas dans une autre, cette modélisation n'est pas suffisante. Par ailleurs, la méthode de nommage des concepts extraits des tags OSM procède à la concaténation de la valeur et de la clé. Si ce mécanisme est pertinent pour l'anglais qui place le complément du nom avant le nom, il l'est moins pour le français dont la syntaxe opère à l'inverse. De plus, la méthode de création des concepts aristotéliens qui représentent l'utilisation conjointe de différents tags autour d'un même objet géographique construit le nom du concept en concaténant les noms des concepts extraits des tags utilisés conjointement (cf. section 6.3.3.2). L'ordre des tags associés à un objet de la base de données OSM n'étant pas significatif, nous avons pris le parti de les ranger par ordre alphanumérique. Cette méthode présente l'avantage d'éviter de construire des concepts doublons qui seraient issus d'objets taggés de la même manière (auxquels le même ensemble de tags est associé) mais dont les tags sont servis dans un ordre différent. En revanche, elle ne fait aucun cas des règles de syntaxe spécifique à la langue dans laquelle elle est utilisée. En ce sens, des travaux doivent être menés pour l'**élaboration d'un système de nommage multilingue des concepts de l'ontologie OF4OSM** pour répondre au besoin d'internationalisation du projet mondial OpenStreetMap.

Suivi sur le temps long

Enfin, et dans la continuité des travaux de [Dahl, 2015a,b] qui visent à établir des mesures de similarité entre les tags OSM sur la base de l'étude des modifications apportées aux objets collectés dans la base de données OSM, il nous semble pertinent de garder une trace des modifications de l'ontologie élaborée à partir de ces tags. En effet, nous avons proposé des méthodes de structuration automatique de la folksonomie OSM qui permettent d'extraire, sur une aire d'étude donnée, voire sur l'ensemble du jeu de données OSM, les objets collectés sur cette zone afin de structurer les tags qui leur sont associés en une ontologie. En l'état, notre approche n'est pas en mesure de proposer un suivi des modifications de cette ontologie mais uniquement d'en générer différentes versions, à différents instants. En vue d'atteindre cet objectif, il est nécessaire de proposer des **méthodes d'intégration de nouveaux concepts et de modification/suppression de concepts existants**. Ces méthodes devront permettre d'enregistrer, sur le temps long, les évolutions de l'ontologie OF4OSM et, *a fortiori*, celles de la folksonomie OSM. Cette démarche pourra être rapprochée des travaux autour du traçage des jeux de données spatio-temporels [Harbelot et al., 2015; Ruas et al., 2013; Plumejeaud, 2011].

Bibliographie

- Arsanjani, J. J., Zipf, A., Mooney, P., and Helbich, M. (2015). *OpenStreetMap in GIScience*. Lecture Notes in Geoinformation and Cartography. Springer International Publishing, Cham.
- ASIS&T (2015). Linked Data and The Charm of Weak Semantics. *Bulletin of the American Society for Information Science and Technology*, 41(4).
- Auer, S., Lehmann, J., and Hellmann, S. (2009). LinkedGeoData : Adding a spatial dimension to the Web of data. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, volume 5823 LNCS, pages 731–746.
- Baglatzi, A., Kokla, M., and Kavouras, M. (2012). Semantifying OpenStreetMap. In *The 11th International Semantic Web Conference*, number Iswc, pages 39–48.
- Ballatore, A., Bertolotto, M., and Wilson, D. C. (2013). Geographic knowledge extraction and semantic similarity in OpenStreetMap. *Knowledge and Information Systems*, 37(1) :61–81.
- Bégin, D., Devillers, R., and Roche, S. (2013). Assessing Volunteered Geographic Information (VGI) Quality Based On Contributors' Mapping Behaviours. *ISPRS, XL-2/W1(June)* :149–154.
- Bishr, M. (2011). *Trust & reputation models for human sensor observations*. PhD thesis, University of Muenster.
- Bishr, M. and Kuhn, W. (2007). Geospatial information bottom-up : A matter of trust and semantics. *The European information society*.
- Bruns, A. (2006). Towards produsage : Futures for User-Led Content Production. In Sudweeks, F., Hrachovec, H., and Ess, C., editors, *Cultural Attitudes towards Communication and Technology 2006*, number 2006, pages 275–284. Perth : Murdoch University.

- Canavosio-Zuzelski, R., Agouris, P., and Doucette, P. (2013). A Photogrammetric Approach for Assessing Positional Accuracy of OpenStreetMap© Roads. *ISPRS International Journal of Geo-Information*, 2(2).
- Codescu, M., Horsinka, G., Kutz, O., Mossakowski, T., and Rau, R. (2011). Osmonto-an ontology of openstreetmap tags. *State of the map*.
- Coleman, D. J. (2010). Volunteered geographic information in spatial data infrastructure : An early look at opportunities and constraints. In *Global Spatial Data Infrastructure Association Conference*, pages 1–18.
- Coleman, D. J., Georgiadou, Y., Labonte, J., Observation, E., and Canada, N. R. (2009). Volunteered Geographic Information : The Nature and Motivation of Producers. *International Journal of Spatial Data Infrastructures Research*, 4 :332–358.
- Cukier, K. (2013). Mise en données du monde, le déluge numérique.
- Dahl, M. (2015a). Edit patterns for OpenStreetMap amenities. <http://matiasdahl.iki.fi/2015/edit-patterns-for-openstreetmap-amenities>.
- Dahl, M. (2015b). Finding related amenity tags on OpenStreetMap. <http://matiasdahl.iki.fi/2015/finding-related-amenity-tags-on-the-openstreetmap>.
- Dorn, H., Törnros, T., and Zipf, A. (2015). Quality Evaluation of VGI Using Authoritative Data—A Comparison with Land Use Data in Southern Germany. *ISPRS International Journal of Geo-Information*, 4(3).
- Du, H., Alechina, N., Jackson, M., and Hart, G. (2013). Matching Formal and informal Geospatial Ontologies. In Vandenbroucke, D., Bucher, B., and Crompvoets, J., editors, *Geographic Information Science at the Heart of Europe*, Lecture Notes in Geoinformation and Cartography, pages 155–171. Springer International Publishing, Cham.
- Euzenat, J. and Shvaiko, P. (2013). *Ontology Matching*. Springer Berlin Heidelberg, Berlin, Heidelberg.
- Exel, M. V. and Dias, E. (2010). Towards A Methodology For Trust Stratification in VGI. In *VGI pre-conference at AGG*, pages 2–5.
- Gangemi, A. (2009). What’s in a schema? A Formal Metamodel for ECG and FrameNet. In Huang, C.-r., Calzolari, N., Gangemi, A., Lenzi, A., Oltramari, A., and Prevot, L., editors, *Ontology and the Lexicon*, pages 144–182. Cambridge University Press, Cambridge.

- Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., and Schneider, L. (2002). *Sweetening Ontologies with DOLCE*, volume 24, pages 166–181. Springer Berlin Heidelberg.
- Glimm, B., Horrocks, I., Motik, B., Stoilos, G., and Wang, Z. (2014). Hermit : An owl 2 reasoner. *Journal of Automated Reasoning*, 53(3) :245–269.
- Golder, S. a. (2006). Usage patterns of collaborative tagging systems. *Journal of Information Science*, 32(2) :198–208.
- Goodchild, M. F. (2007). Citizens as Sensors : the World of Volunteered Geography. *GeoJournal*, 69(4) :211–221.
- 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.
- Harbelot, B., Arenas, H., and Cruz, C. (2015). LC3 : A spatio-temporal and semantic model for knowledge discovery from geospatial datasets. *Journal of Web Semantics*.
- Ho, S. and Rajabifard, A. (2010). Learning from the crowd : The role of volunteered geographic information in realising a spatially enabled society. In *GSDI Conference Proceedings*.
- Hombiat, A., Gensel, J., and Villanova-Oliver, M. (2015). Un nouveau méta-modèle pour rapprocher la folksonomie et l’ontologie d’OSM. In *Colloque International de Géomatique et d’Analyse Spatiale (SAGEO 2015)*, Hammamet, Tunisia.
- Hombiat, A., Villanova-Oliver, M., and Gensel, J. (2016). Un métamodèle pour structurer la folksonomie d’osm en une nouvelle ontologie. *Revue Internationale de Géomatique*, 26(4) :425–444.
- Jilani, M., Corcoran, P., and Bertolotto, M. (2014). Automated highway tag assessment of openstreetmap road networks. In *22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems, ACM SIGSPATIAL GIS 2014*, volume 04-07-Nove.
- Johnson, P. A. and Sieber, R. E. (2013). Situating the Adoption of VGI by Government. In *Crowdsourcing Geographic Knowledge*, pages 65–81. Springer Netherlands, Dordrecht.
- Kesler, C., Theodore, R., and Groot, A. D. (2013). *Trust as a Proxy Measure for the Quality of Volunteered Geographic Information in the Case of OpenStreetMap*, pages 21–37. Springer International Publishing.

- Kuhn, W. (2005). Geospatial Semantics : Why , of What , and How ? Introduction : Why Semantics ? *Journal on Data Semantics III*, 3534 :1–24.
- Loenen, B. V. (2012). Quest for a global standard for geo-data licenses. In *Spatially enabling government, industry and citizens : research and development perspectives*, chapter 3, pages 39–55. GSDI Association Press, Needham.
- Lohmann, S., Díaz, P., and Aedo, I. (2011). MUTO. In *Proceedings of the 7th International Conference on Semantic Systems - I-Semantics '11*, pages 95–104.
- McCall, M. K. and Minang, P. A. (2005). Assessing participatory GIS for community-based natural resource management : Claiming community forests in Cameroon. *Geographical Journal*, 171(4) :340–356.
- Mooney, P. and Corcoran, P. (2012). Characteristics of Heavily Edited Objects in OpenStreetMap. *Future Internet*, 4(4) :285–305.
- Neis, P., Zielstra, D., and Zipf, A. (2011). The Street Network Evolution of Crowdsourced Maps : OpenStreetMap in Germany 2007–2011. *Future Internet*, 4(1) :1–21.
- Oldenburg, R. (1989). *The Great Good Place : Cafes, Coffee Shops, Bookstores, Bars, Hair Salons, and Other Hangouts at the Heart of a Community*. New York : Marlowe.
- O’Reilly, T. (2005). What Is Web 2.0 Design Patterns and Business Models for the Next Generation of Software. <http://www.oreilly.com/pub/a/web2/archive/what-is-web-20.html>.
- Plumejeaud, C. (2011). *Modèles et méthodes pour l’information spatio-temporelle évolutive*. PhD thesis, Université de Grenoble.
- Poole, D., Smyth, C., and Sharma, R. (2008). Semantic science : Ontologies, data and probabilistic theories. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5327 LNAI :26–40.
- Raymond, E. (1999). The cathedral and the bazaar. *Knowledge, Technology & Policy*, 12(3) :23–49.
- Rehrl, K., Gröechnig, S., Hochmair, H., Leitinger, S., Steinmann, R., and Wagner, A. (2013). A Conceptual Model for Analyzing Contribution Patterns in the Context of VGI. In *Progress in Location-Based Services*, pages 373–388.

- Ruas, A., Plumejeaud-Perreau, C., NAHASSIA, L., Eric, G., Olteanu, A.-M., VOULOIR, M. C., and MOTTE, C. (2013). GéoPeuple : création et analyse de données topographiques et démographiques sur 200 ans. *Revue du comité français de cartographie*, (218) :17–29.
- Salzano, G. (2013). Réutilisation de données publiques Une étude comparative entre quatre pays. In *INFORSID 2013*, volume X, pages 371–386, Paris.
- Sébastien, M., Abadie, N., Aussenac-Gilles, N., Bessagnet, M.-N., Kamel, M., Kergosien, E., Reynaud, C., Safar, B., and Sallaberry, C. (2011). Analyses linguistiques et techniques d’alignement pour créer et enrichir une ontologie topographique. *Revue Internationale de Géomatique*, 21(2) :155–180.
- Sehra, S. S. (2014). Assessing the Topological Consistency of Crowdsourced OpenStreetMap Data. *Human Computation*, 1(2) :265–280.
- Shirky, C. (2005). Ontology is overrated : categories, links, and tags. <http://www.citeulike.org/group/2924/article/221521>.
- Silvertown, J. (2009). A new dawn for citizen science. *Trends in Ecology and Evolution*, 24(9) :467–471.
- Sirin, E., Parsia, B., Grau, B. C., Kalyanpur, A., and Katz, Y. (2007). Pellet : A practical owl-dl reasoner. *Web Semantics : Science, Services and Agents on the World Wide Web*, 5(2) :51 – 53. Software Engineering and the Semantic Web.
- Sowa, J. F. (2000). *Knowledge Representation : Logical, Philosophical and Computational Foundations*. Brooks/Cole Publishing Co., Pacific Grove, CA, USA.
- Togia, T. (2005). Ontology vs. Folksonomy. <http://www.citeulike.org/group/626/article/421963>.
- Tsarkov, D. and Horrocks, I. (2006). Fact++ description logic reasoner : System description. In *Proceedings of the Third International Joint Conference on Automated Reasoning, IJCAR’06*, pages 292–297, Berlin, Heidelberg. Springer-Verlag.
- Vandecasteele, A. and Devillers, R. (2015). *Improving Volunteered Geographic Information Quality Using a Tag Recommender System : The Case of OpenStreetMap*, chapter Improving, pages 59–80. Springer International Publishing, Cham.
- Wal, T. V. (2005). Folksonomy Definition and Wikipedia. <http://www.vanderwal.net/random/entrysel.php?blog=1750>.
- Weidmann, N. B., Benitez-Baleato, S., Hunziker, P., Glatz, E., and Dimitropoulos, X. (2016). Digital discrimination : Political bias in internet service provision across ethnic groups. *Science*, 353(6304) :1151–1155.

- Weller, K. (2007). Folksonomies and ontologies : two new players in indexing and knowledge representation. *Online Information 2007*, pages 108–115.
- Yang, T. A., Kim, D. J., Dhalwani, V., and Vu, T. K. (2008). The 8C framework as a reference model for collaborative value webs in the context of Web 2.0. In *Proceedings of the Annual Hawaii International Conference on System Sciences*, page 319.
- Zielstra, D. and Zipf, A. (2010). A comparative study of proprietary geodata and volunteered geographic information for Germany. *13th AGILE International Conference on Geographic Information Science 2010 Guimarães, Portugal*, 1 :1–15.

Annexes

Annexe A

Ontologie MUTO

(ontologie unifiée de l'activité de
tagging)

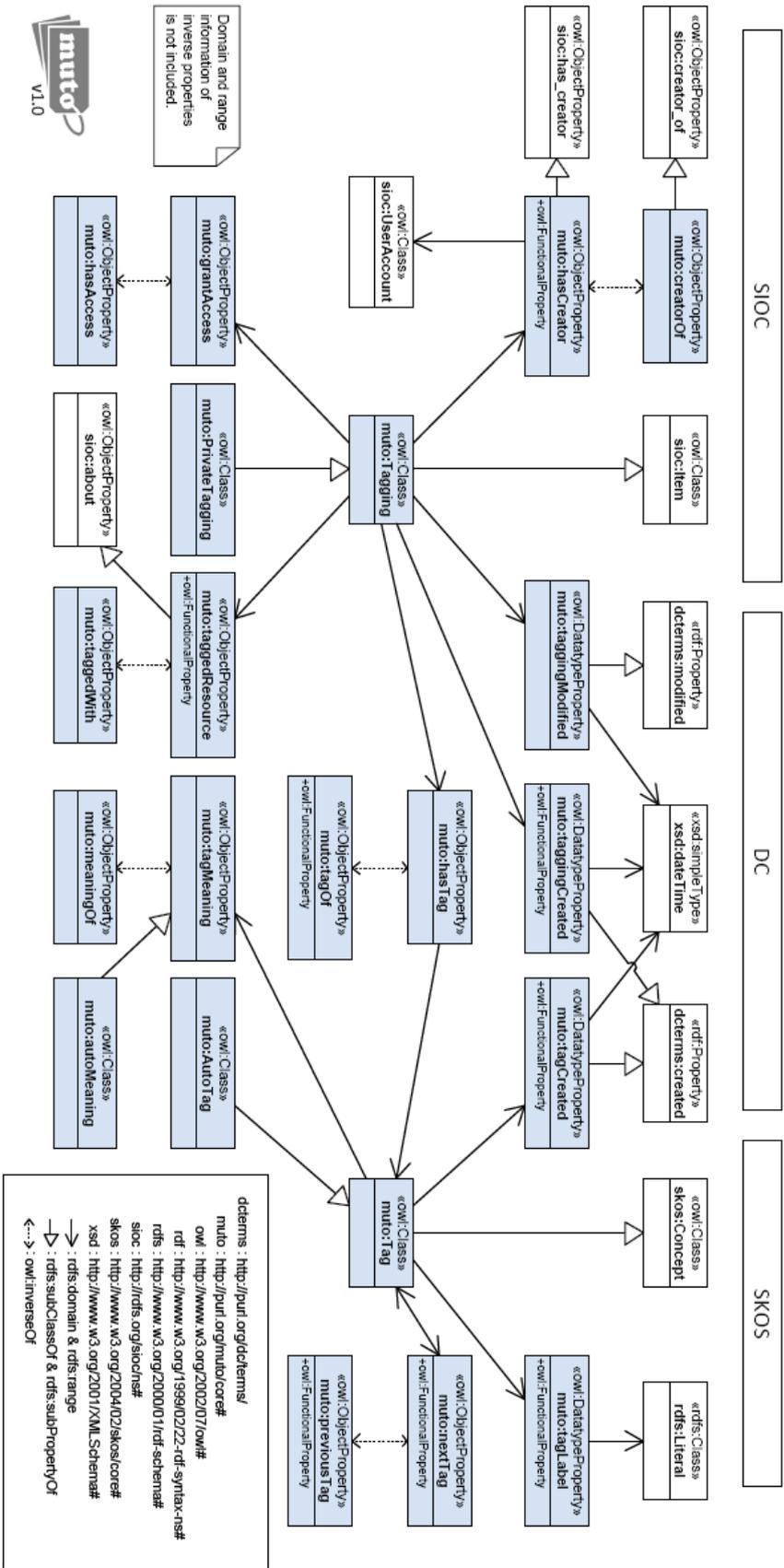


FIGURE A.1 – ontologie unifiée de l'activité de tagging (Modular Unified Tagging Ontology, MUTO)

Annexe B

Langage OWL
(*Ontology Web Language*)

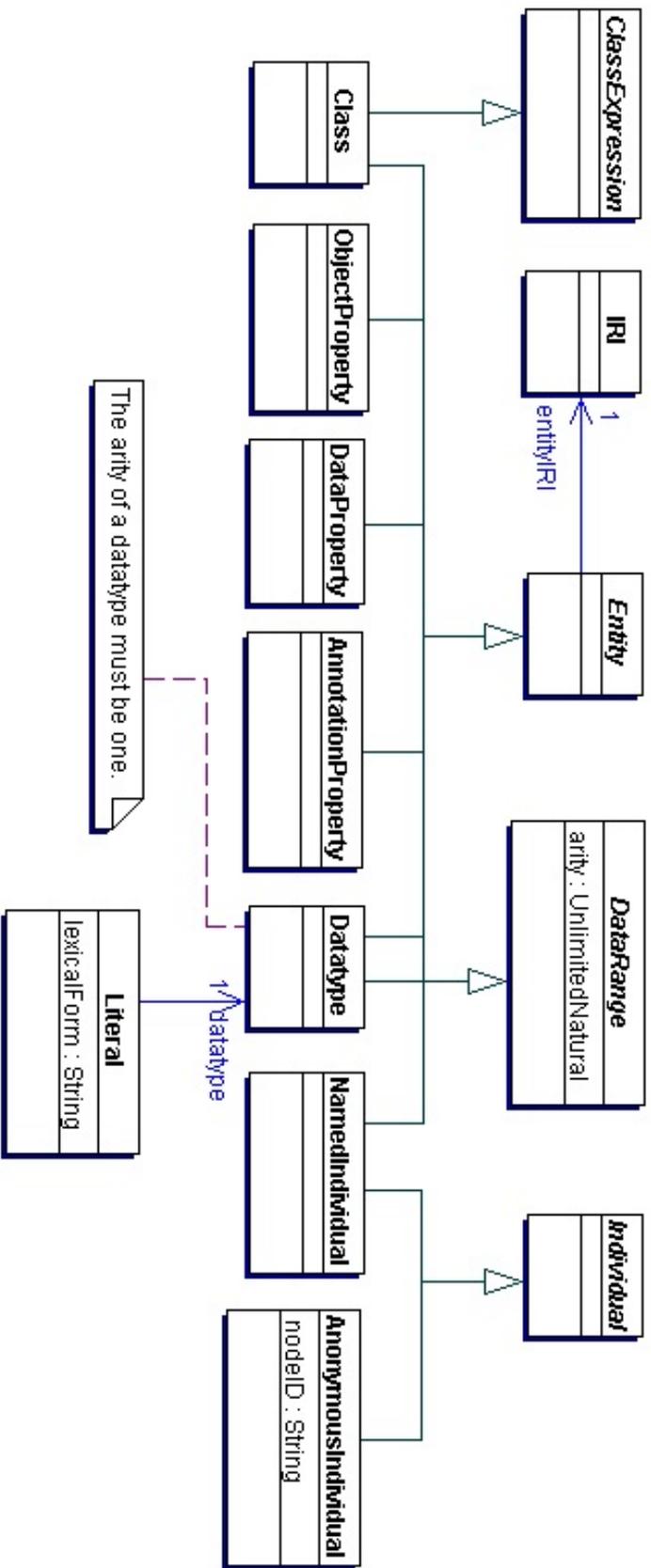


FIGURE B.1 – *Ontology Web Language (OWL)*

Annexe C

Valeurs multiples associées à la clé amenity

Valeurs multiples associées à la clé <code>amenity</code>	Nb occurrences
<code>parking ;fuel</code>	1387
<code>waste _basket ;recycling</code>	398
<code>school ;place _of _worship / place _of _worship ;school</code>	89 (80 + 9)
<code>bar ;restaurant / restaurant ;bar</code>	110 (76 + 34)
<code>restaurant ;cafe / cafe ;restaurant</code>	82 (43 + 39)
<code>restaurant ;fast _food</code>	34
<code>bench ;waste _basket / waste _basket ;bench</code>	43 (32 + 11)
<code>pub ;restaurant</code>	27
<code>shelter ;bench</code>	23
<code>vending _machine ;waste _basket / waste _basket ;vending _machine / vending _machine ; waste _basket</code>	40 (23 + 9 + 8)
<code>toilets ;shower</code>	23
<code>fuel ;car _wash</code>	20
<code>parking ;restaurant ;fuel</code>	18
<code>cafe ;bar</code>	16
<code>restaurant ;pub</code>	15
<code>university ;hospital</code>	15
<code>atm ;bank / bank ; atm / bank ;atm</code>	32 (13 + 10 + 9)
<code>fuel ;food _court</code>	13
<code>cafe ;fast _food</code>	12
<code>school ;kindergarten</code>	11
<code>doctors ;dentist</code>	10
<code>fast _food ;cafe</code>	10
<code>kindergarten ;school</code>	10
<code>bar ;cafe</code>	10
<code>shower ;toilets</code>	9
<code>shelter ;bicycle _parking ;bicycle _rental</code>	7
<code>fuel ;compressed _air</code>	7
Total	2471

TABLE C.1 – Valeurs multiples associées à la clé `amenity`

Annexe D

Ontologie AO

(Ontologie Aristotélicienne)

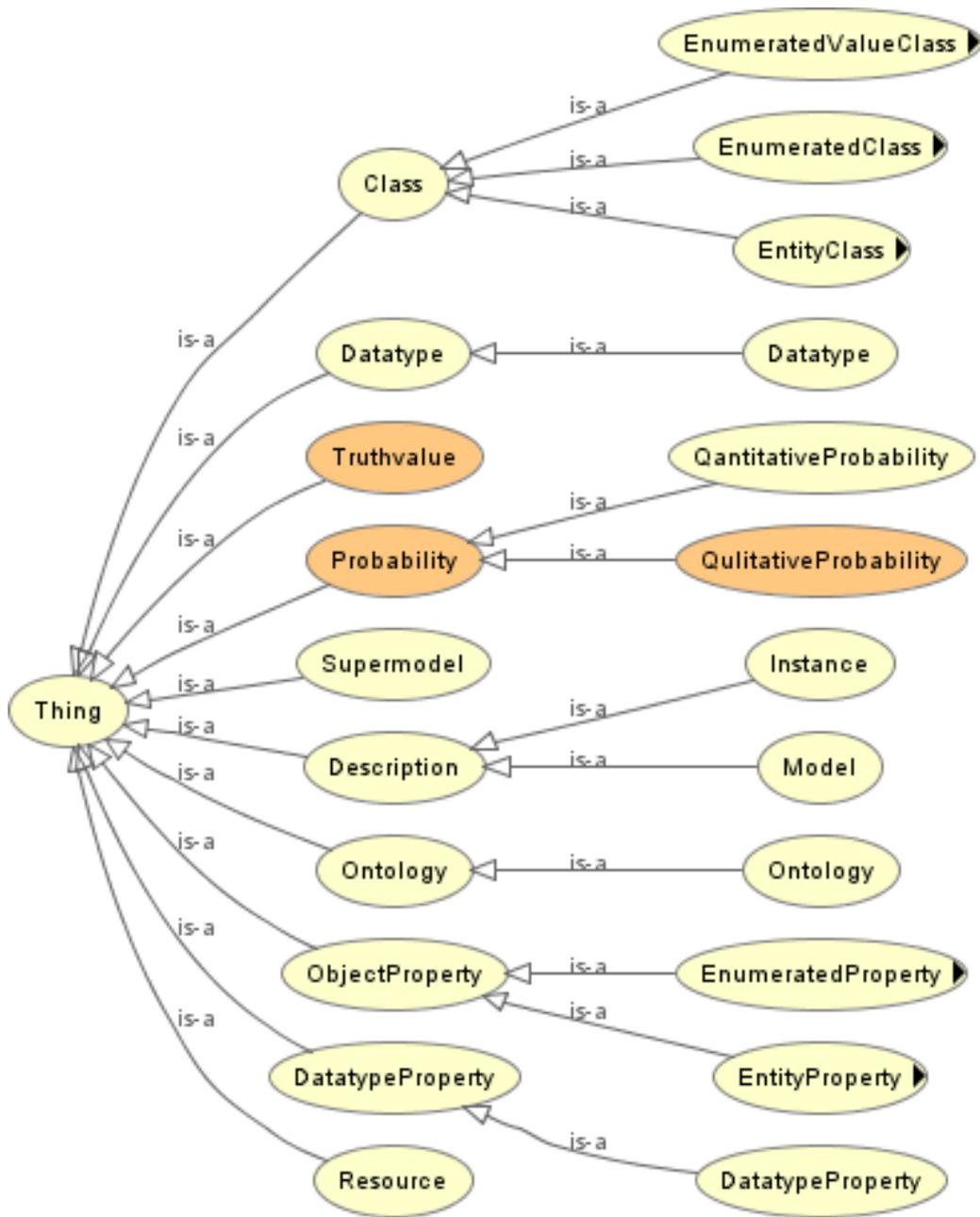


FIGURE D.1 – Ontologie Aristotélienne (*Aristotelian Ontology, AO*)

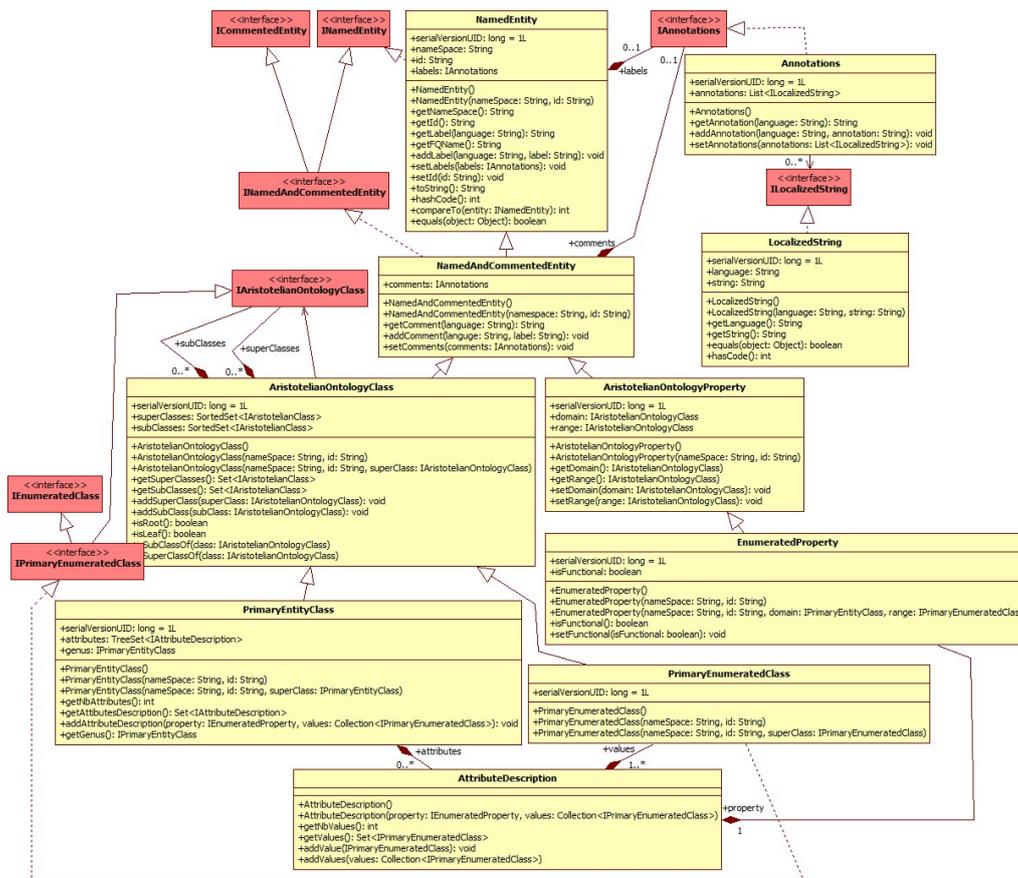


FIGURE D.2 – Diagramme de classe de l’API pour Editeur de Classes Aristotéliennes (*Aristotelian Class Editor, ACE*)

Annexe E

Extrait du code HTML de la page *Web Map Features* sur le wiki OSM

```
1 <h3>Amenity</h3>
2 <p>Used to map facilities used by visitors and residents.</p>
3 <table class="wikitable">
4   <tr>
5     <th>Key</th>
6     <th>Value</th>
7     <th>Element</th>
8     <th>Comment</th>
9     <th>Rendering</th>
10    <th>Photo</th>
11  </tr>
12  <tr>
13    <th colspan="7">
14      <h4>Sustenance</h4>
15    </th>
16  </tr>
17  <tr id="amenity-bar">
18    <td><a href="/wiki/Key:amenity">amenity</a></td>
19    <td><a href="/wiki/Tag:amenity%3Dbar">bar</a></td>
20    <td>
21      <a href="/wiki/Elements#Node"></a>
22      <a href="/wiki/Elements#Area"></a>
23    </td>
24    <td><b>Bar</b> is a purpose-built commercial establishment
25      that sells alcoholic drinks.</td>
26    <td>
27      <a href="/wiki/File:bar-icon.png"></a>
28    </td>
29    <td>
30      <a href="/wiki/File:bar.png"></a>
```

```

31 </td>
32 </tr>
33 <tr id="amenity-bbq">
34 <td><a href="/wiki/Key:amenity">amenity</a></td>
35 <td><a href="/wiki/Tag:amenity%3Dbbq">bbq</a> </td>
36 <td>
37 <a href="/wiki/Elements#Node"></a>
38 </td>
39 <td><b>BBQ</b> or <b>Barbecue</b> is a permanently built
40 grill for cooking food</td>
41 <td></td>
42 <td><a href="/wiki/File:bbq.png"></a></td>
43 </tr>
44 <tr id="amenity-biergarten">
45 <td><a href="/wiki/Key:amenity">amenity</a></td>
46 <td><a href="/wiki/Tag:amenity%3Dbiergarten">biergarten</a></td>
47 <td>
48 <a href="/wiki/Elements#Node"></a>
49 <a href="/wiki/Elements#Area"></a>
50 </td>
51 <td><b>Biergarten</b> is an open-air area where alcoholic beverages
52 along with food is prepared and served.</td>
53 <td>
54 <a href="/wiki/File:biergarten-icon.png"></a>
55 </td>
56 <td>
57 <a href="/wiki/File:biergarten.png"></a>
58 </td>
59 </tr>
60 ...

```