

# OF4OSM

## Une nouvelle base de connaissances extraite des tags OpenStreetMap

---

thèse de doctorat présentée par  
**Anthony Hombiat**

sous la direction de  
**M. Jérôme Gensel et Mme Marlène Villanova-Oliver**

**Équipe STeamer**  
**Laboratoire d'Informatique de Grenoble**

16 janvier 2017

# Plan

## Contexte et problématique

### I. Méthodes de structuration des tags OSM existantes

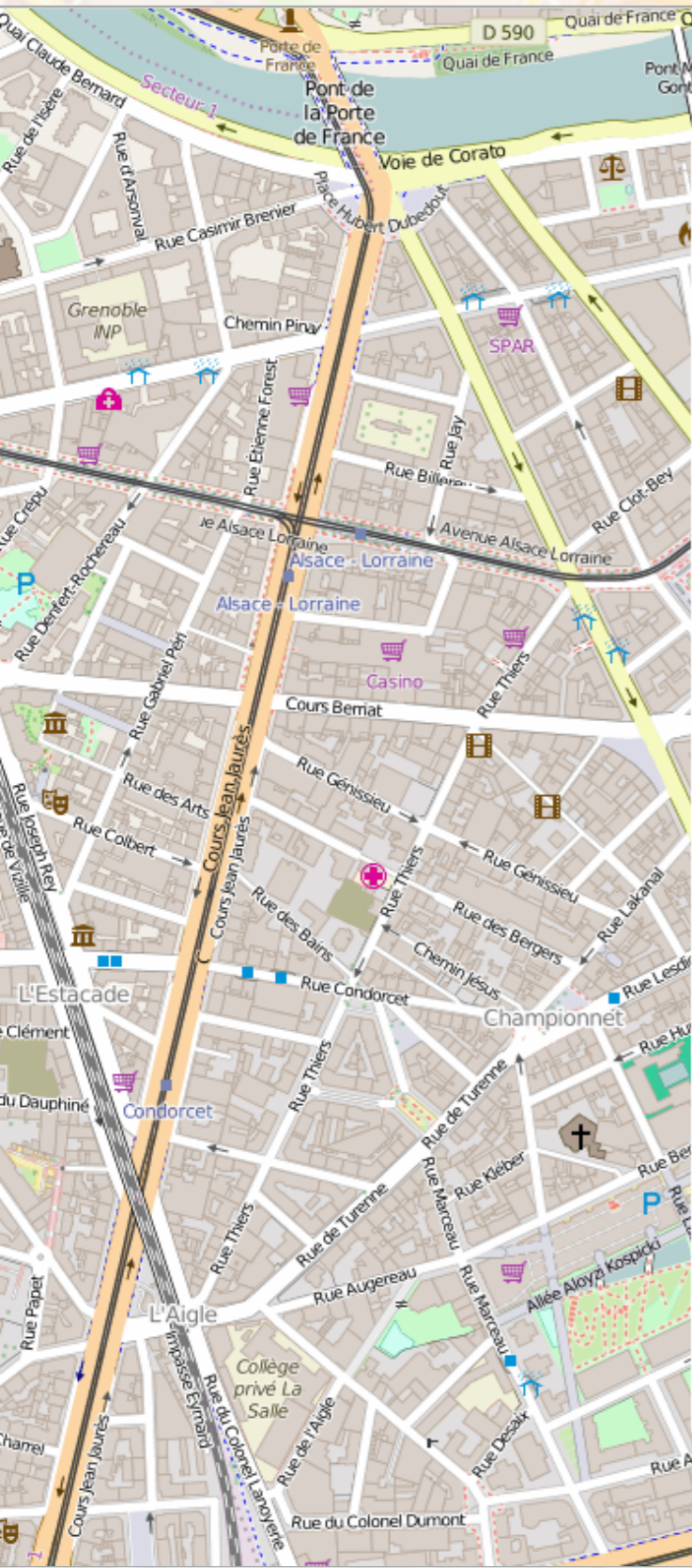
- ▶ Wiki OSM
- ▶ Linked Geo Data (LGD)
- ▶ OSMonto
- ▶ Game With A Purpose (GWAP)
- ▶ OSM Semantic Network (OSN)

### II. OF4OSM : nouvelle base de connaissances OSM

- ▶ Représentation d'un tag spécifique à OSM
- ▶ Typologie des tags OSM
- ▶ Génération de OF4OSM

## Bilan et perspectives

# Contexte et problématique



## Relation : Stade des Alpes (80400)

### Corrections building Stade des Alpes

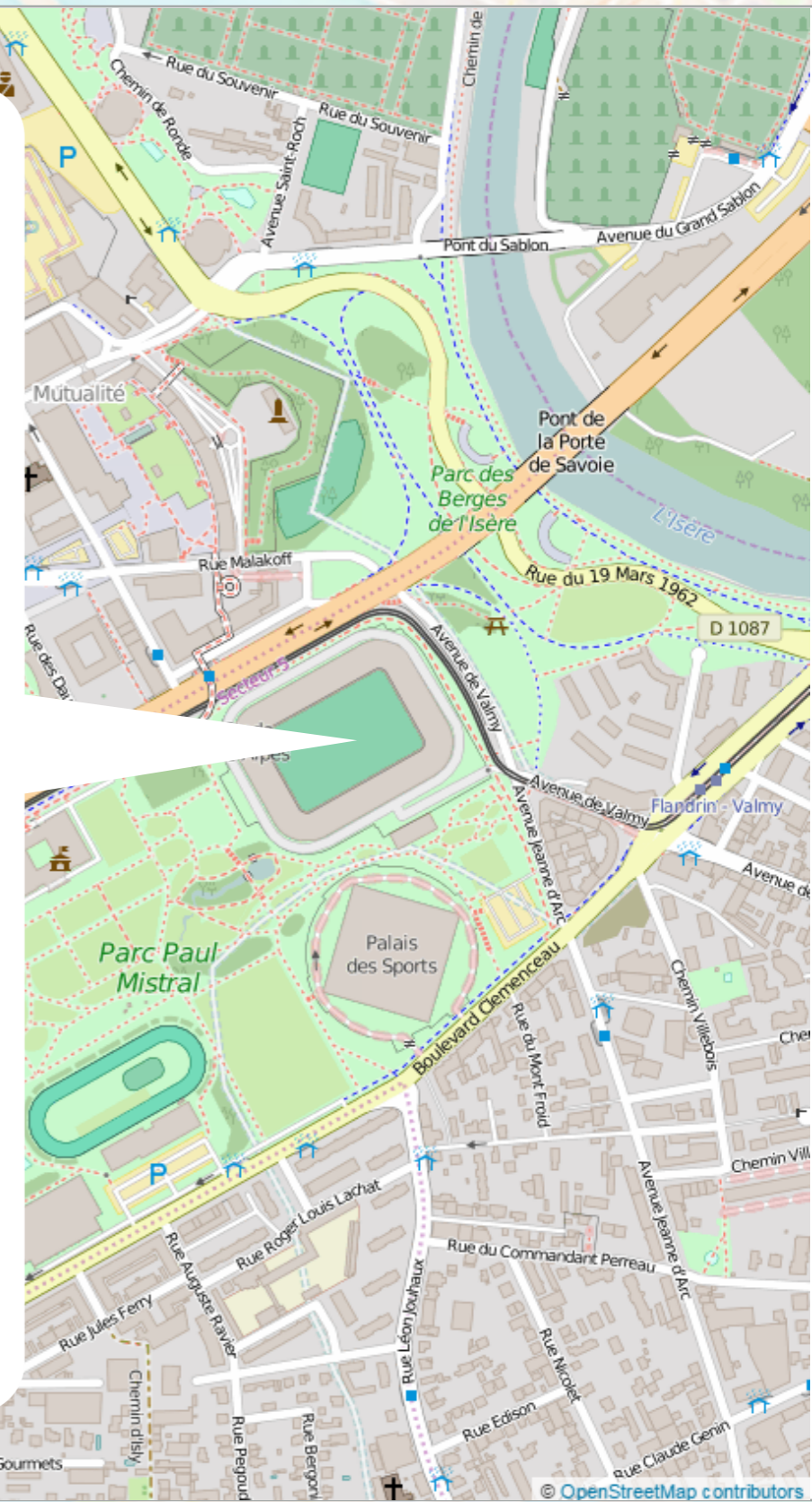
Modifié il y a environ 3 ans par [redacted]  
Version #4 · Groupe de modifications #19233634

### Attributs

building	yes
height	23
leisure	stadium
name	Stade des Alpes
sport	soccer
type	multipolygon
website	<a href="http://www.stadedesalpes.fr/">http://www.stadedesalpes.fr/</a>
wikipedia	<a href="fr:Stade des Alpes">fr:Stade des Alpes</a>

### Membres

- Chemin 25532413 avec le rôle outer
- Chemin 25532414 avec le rôle inner



- ▶ Les **données attributaires** des objets de la BD OSM sont représentées par des **tags**. Ces tags sont sous la forme d'une paire **clé=valeur** en **texte libre**.
- ▶ Avantages :
  - ▶ **Facile** à utiliser pour les contributeurs
  - ▶ Très **évolutif**
- ▶ Inconvénients :
  - ▶ Obstacle à l'**acquisition**
  - ▶ Obstacle à la **restitution**

### Problèmes liés à la catégorisation de contenu par tags :

- ▶ Erreurs typographiques

● `liesure=park`

- ▶ Redondance

● `leisure=swimming_pool`

● `amenity=swimming_pool`

- ▶ Ambiguïté sémantique

● `building=yes`

● `building=castle`

● `access=private`

- ▶ Les tags ne sont pas organisés entre eux

● `amenity=restaurant`



● `cuisine=*`



Comment structurer les tags OpenStreetMap sans réduire la flexibilité de leur utilisation ?

# Plan

## Contexte et problématique

### I. Méthodes de structuration des tags OSM existantes

- ▶ Wiki OSM
- ▶ Linked Geo Data (LGD)
- ▶ OSMonto
- ▶ Game With A Purpose (GWAP)
- ▶ OSM Semantic Network (OSN)

### II. OF4OSM : nouvelle base de connaissances OSM

- ▶ Représentation d'un tag spécifique à OSM
- ▶ Typologie des tags OSM
- ▶ Génération de OF4OSM

## Bilan et perspectives



## ▶ Wiki OSM

- ▶ + 2000 tags listés
- ▶ Pages dédiées aux clés et aux valeurs
- ▶ Description de l'usage des tags
- ▶ Ouvert à toute contribution
- ▶ Processus de validation de nouveaux tags
- ▶ Cycle Contribution/Révocation/Discussion (BRD cycle)



- ▶ **Linked Geo Data (LGD)** [Auer *et al.*, 2009]

- ▶ 3 types de tags :

- ▶ Attributs de **classification**  `historic=castle`

- ▶ Attributs de **description**  `internet_access=wlan`

- ▶ Attributs de **données**  `height=80`

- ▶ Identification des tags via wiki OSM :

<https://wiki.openstreetmap.org/wiki/Tag:amenity=school>

- ▶ Ne considère que les tags du wiki OSM

# Méthodes de structuration des tags OSM

- ▶ OSMonto [Codescu *et al.*, 2011]

- ▶ Subsumption **clé > valeur**

- ▶ Préfixes **k\_** et **v\_** pour désambiguïser

● amenity=school

● school=primary

- ▶ Relations de **dépendance** entre tags

● amenity=restaurant

● cuisine=french

● cuisine=libanese

● cuisine=chinese

- ▶ Tags du wiki OSM + > 100 occurrences dans la BD OSM

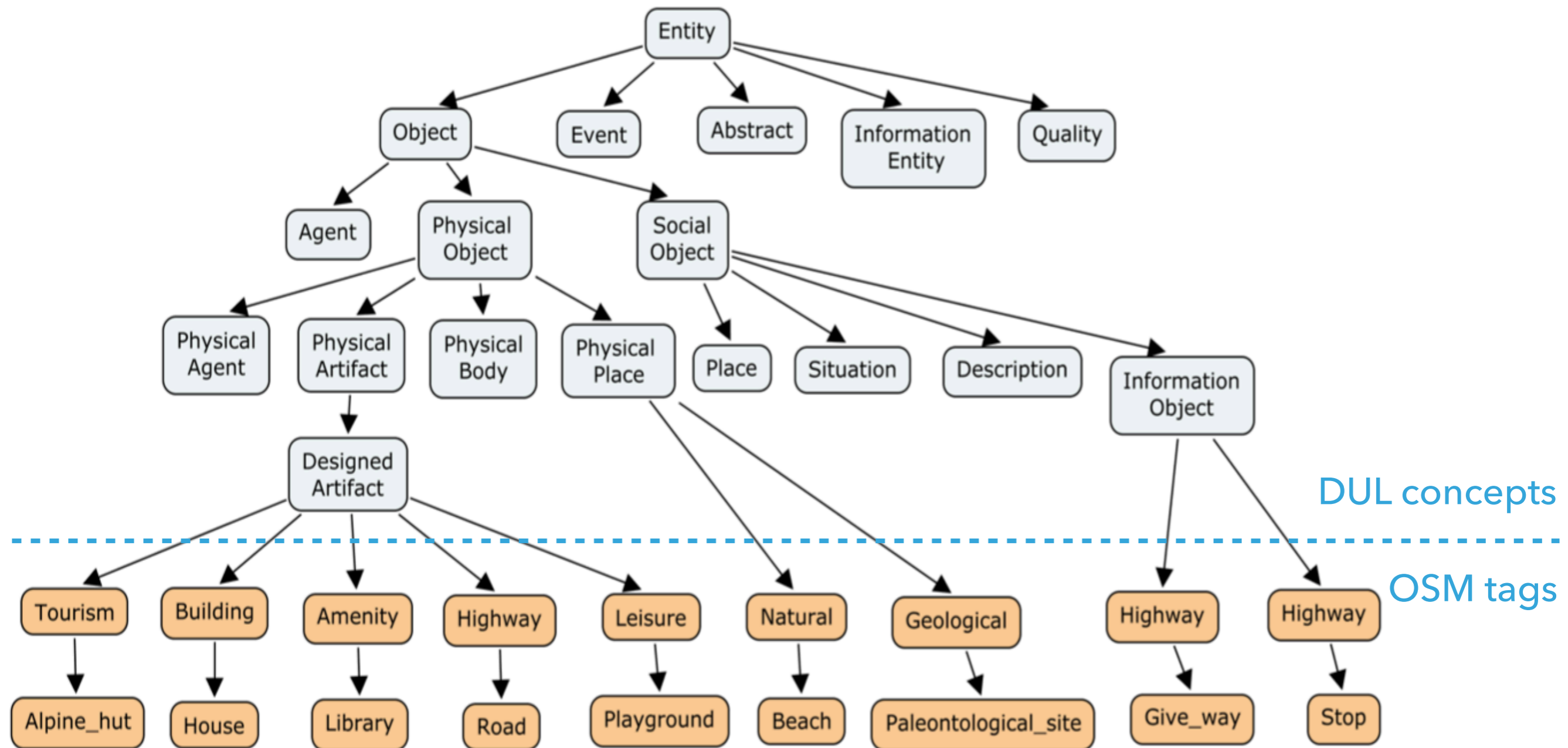
- k\_historic
- k\_landuse
- k\_leisure
  - v\_arena
  - v\_common
  - v\_dance
  - v\_dog\_park
  - v\_fishing
  - v\_fitness\_station
  - v\_garden
  - v\_golf\_course
  - v\_ice\_rink
  - v\_marina
  - v\_miniature\_golf
  - v\_nature\_reserve
  - v\_park
  - v\_picnic\_table
  - v\_pitch
  - v\_playground
  - v\_recreation\_ground
  - v\_slipway
  - v\_sports\_centre
  - v\_stadium
  - v\_swimming\_pool
  - v\_track
  - v\_water\_park
- k\_man\_made
- k\_military
- k\_natural

- ▶ **Game With A Purpose (GWAP)** [Baglatzi *et al.*, 2012]
  - ▶ Modèle de **haut niveau** DUL (Dolce + DnS Ultra Lite)
  - ▶ **1 concept DUL = 1 question contributeur OSM**

1. Is it a (physical) object like a river or a stadium?
2. Is it a material, like sand or mud?
3. Is it a boundary of an area like an electoral division?
4. Does it imply some kind of action like swimming or dancing?
5. Can you observe and measure it?
6. Does it have a location?
7. ...

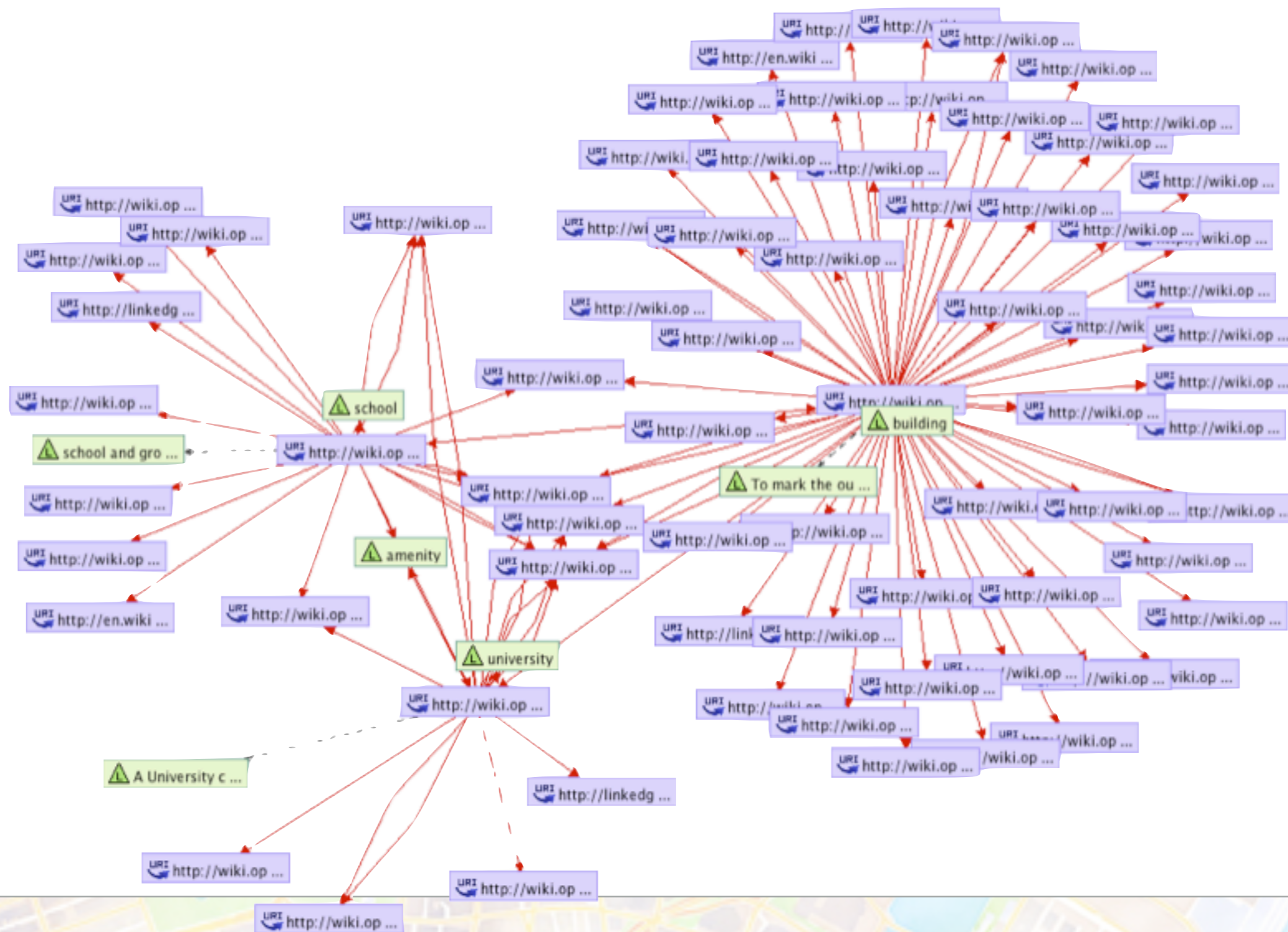
**Table 1.** Sample questions for the alignment process

## ► Game With A Purpose (GWAP) [Baglatzi et al., 2012]



- ▶ OSM Semantic Network (OSN) [Ballatore *et al.*, 2013]
  - ▶ Mesures de **similarité** entre tags
  - ▶ **Web crawler** wiki OSM (liens entrants et sortants)
  - ▶ **Concepts et individus** de l'ontologie LGD

## ► OSM Semantic Network (OSN) [Ballatore et al., 2013]



# Méthodes de structuration des tags OSM

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

Comparaison des méthodes de structuration des tags OSM

# Plan

## Contexte et problématique

### I. Méthodes de structuration des tags OSM existantes

- ▶ Wiki OSM
- ▶ Linked Geo Data (LGD)
- ▶ OSMonto
- ▶ Game With A Purpose (GWAP)
- ▶ OSM Semantic Network (OSN)

### II. OF4OSM : nouvelle base de connaissances OSM

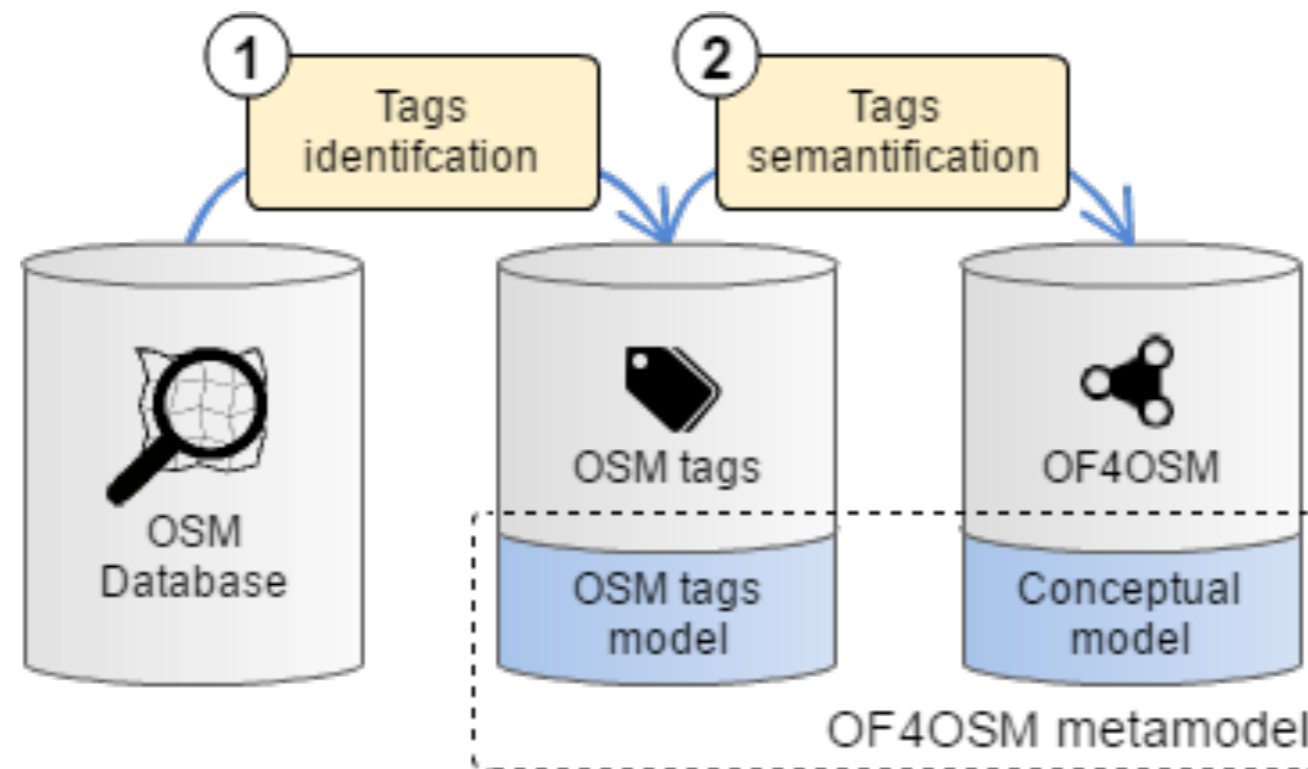
- ▶ Représentation d'un tag spécifique à OSM
- ▶ Typologie des tags OSM
- ▶ Génération de OF4OSM

## Bilan et perspectives



## Construction de la base de connaissances OF4OSM en 2 étapes :

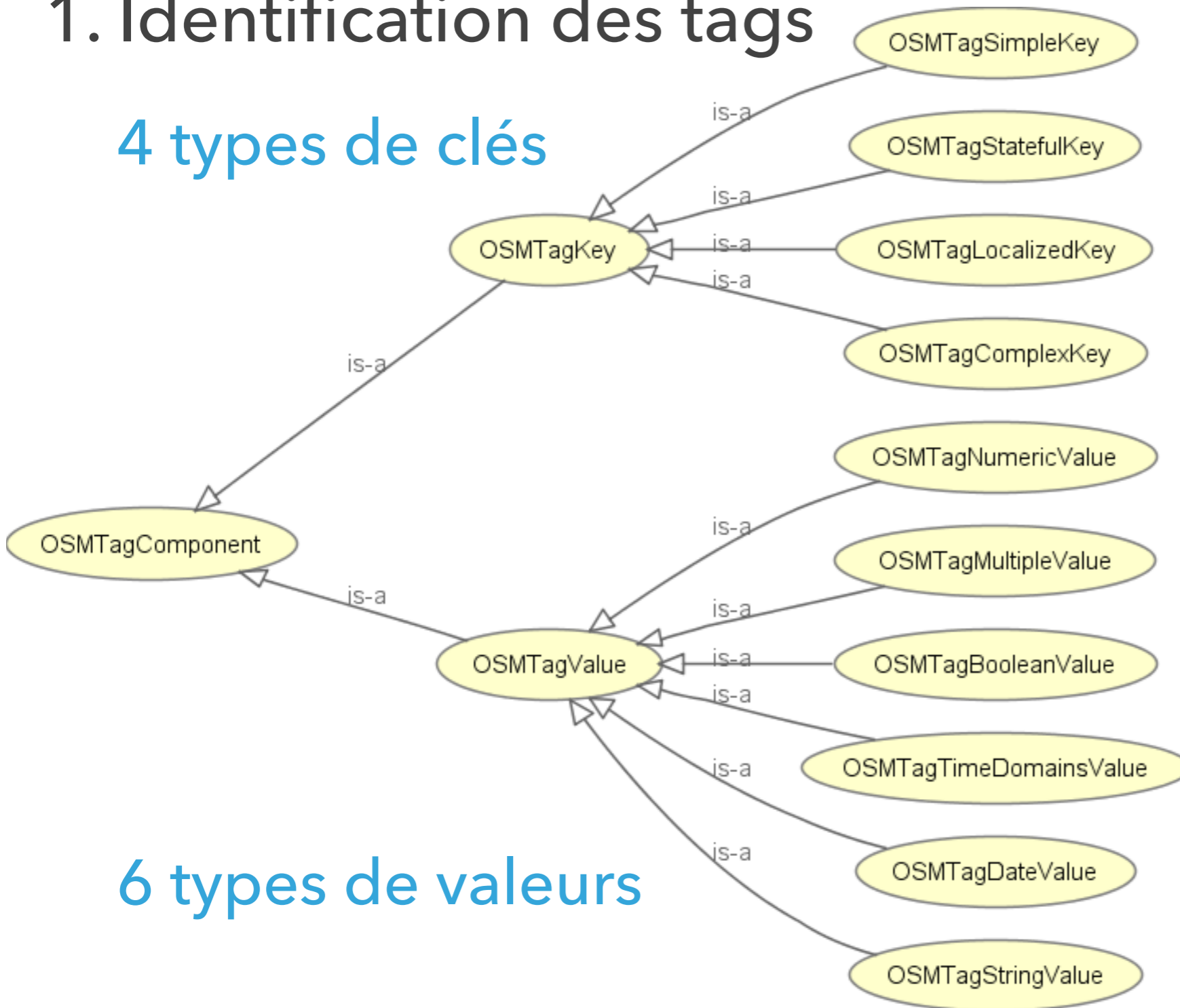
1. **Identification** des tags OSM
2. **Sémantification** des tags OSM



## 1. Identification des tags

4 types de clés

6 types de valeurs



sport

amenity:abandoned

toilets:wheelchairs

● sport=soccer

● amenity:abandoned=theater

● toilets:wheelchairs=true

true

soccer

theater

## 1. Identification des tags

	TYPE	CLÉ	VALEUR	EXEMPLE
<i>Category Tag</i>	IOSMSimpleCategoryTag	IOSMTagSimpleKey	IOSMTagStringValue	amenity=school
	IOSMMultipleCategoryTag	IOSMTagSimpleKey	IOSMTagMultipleValue	amenity=cafe;library
	IOSMStatefulCategoryTag	IOSMTagStatefulKey	IOSMTagStringValue	amenity:abandoned=theater
<i>Property Tag</i>	IOSMSimpleKeyStringValuePropertyTag	IOSMTagSimpleKey	IOSMTagStringValue	name=London
	IOSMComplexKeyStringValuePropertyTag	IOSMTagComplexKey	IOSMTagStringValue	addr:street=JeanJaures
	IOSMLocalizedKeyStringValuePropertyTag	IOSMTagLocalizedKey	IOSMTagStringValue	name:fr=London
	IOSMSimpleKeyBooleanValuePropertyTag	IOSMTagSimpleKey	IOSMTagBooleanValue	drive_in=false
	IOSMComplexKeyBooleanValuePropertyTag	IOSMTagComplexKey	IOSMTagBooleanValue	toilets:wheelchairs=true
	IOSMSimpleKeyNumericValuePropertyTag	IOSMTagSimpleKey	IOSMTagNumericValue	height=80
	IOSMComplexKeyNumericValuePropertyTag	IOSMTagComplexKey	IOSMTagNumericValue	addr:housenumber=123
	IOSMSimpleKeyDateValuePropertyTag	IOSMTagSimpleKey	IOSMTagDateValue	start_date=2000-01-01
	IOSMComplexKeyDateValuePropertyTag	IOSMTagComplexKey	IOSMTagDateValue	source:date=2000-01-01
	IOSMTimeDomainsValuePropertyTag	IOSMTagSimpleKey	IOSMTagTimeDomainsValue	opening_hours=Mo-Fr08:30-16:30
IOSMMultipleValuePropertyTag	IOSMTagSimpleKey	IOSMTagMultipleValue	service=repair;drive-through	

## Typologie des tags OSM

## 2. Sémantification des tags OSM en 3 étapes :

1. Hiérarchie clé > tag
2. Concepts de haut niveau
3. Combinaison de tags

### ► Exemple :



historic = castle

tourism = hotel

amenity = restaurant

Le Château de la Commanderie à Eybens en Isère

<http://www.openstreetmap.org/way/66383517>

# OF4OSM : une nouvelle base de connaissances OSM

1. Hiérarchie clé > tag
2. Concepts de haut niveau
3. Combinaison de tags

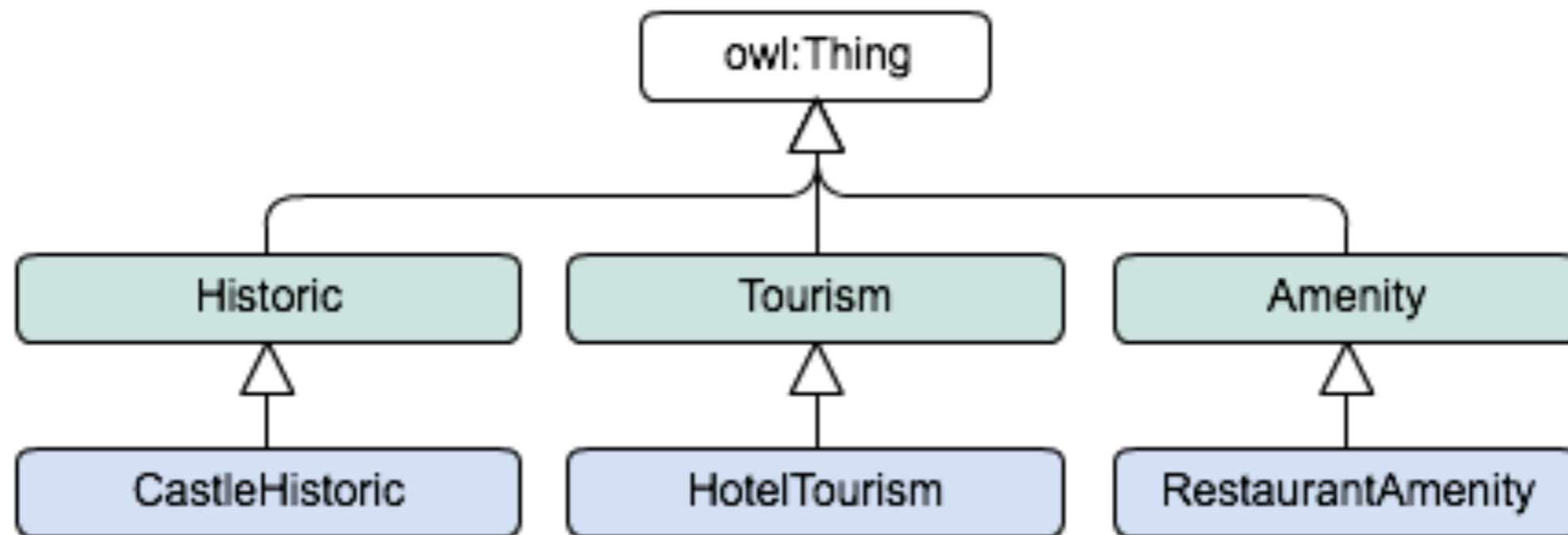


historic = castle

tourism = hotel

amenity = restaurant

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



# OF4OSM : une nouvelle base de connaissances OSM

1. Hiérarchie clé > tag
2. Concepts de haut niveau
3. Combinaison de tags

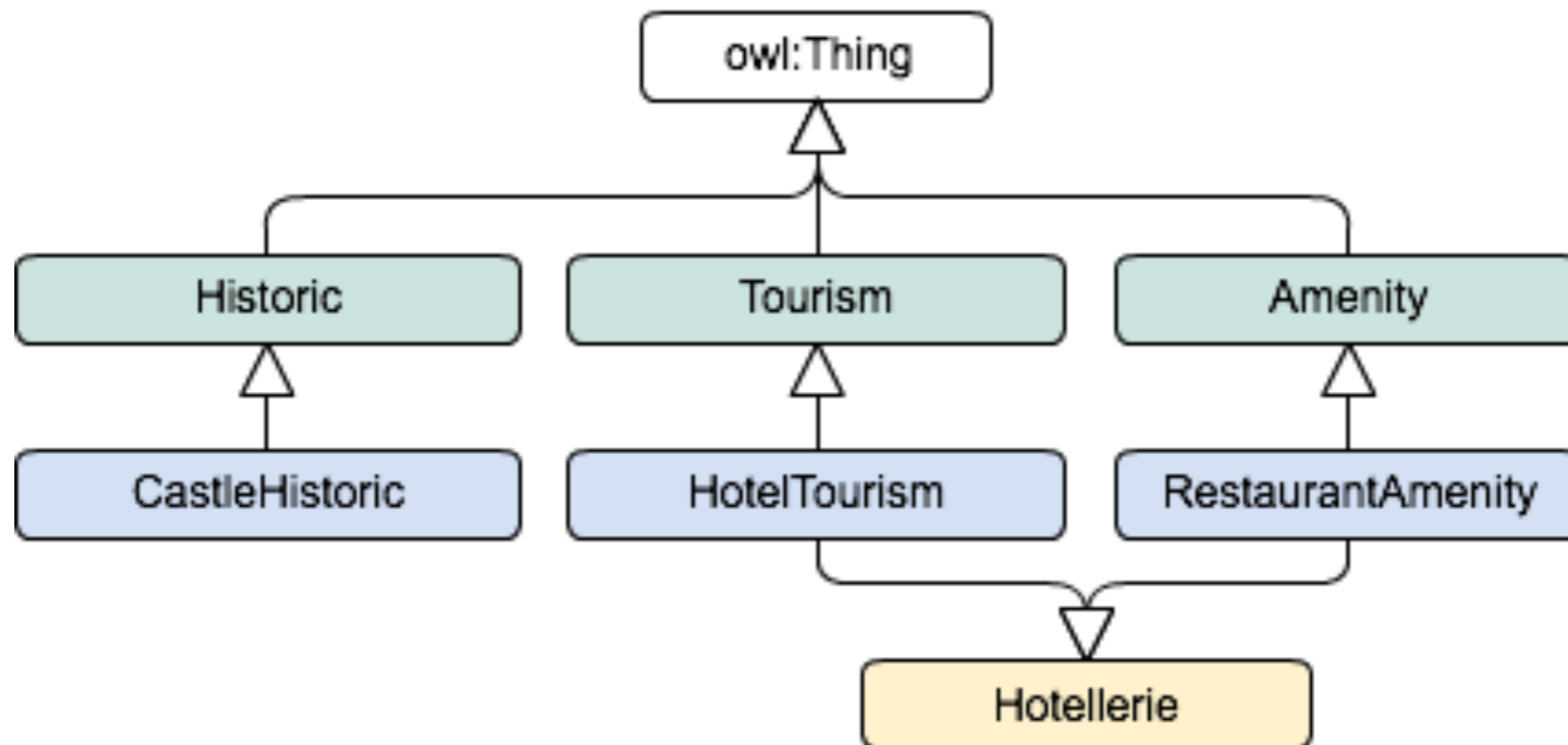


historic = castle

tourism = hotel

amenity = restaurant

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



# OF4OSM : une nouvelle base de connaissances OSM

1. Hiérarchie clé > tag
2. Concepts de haut niveau
3. Combinaison de tags

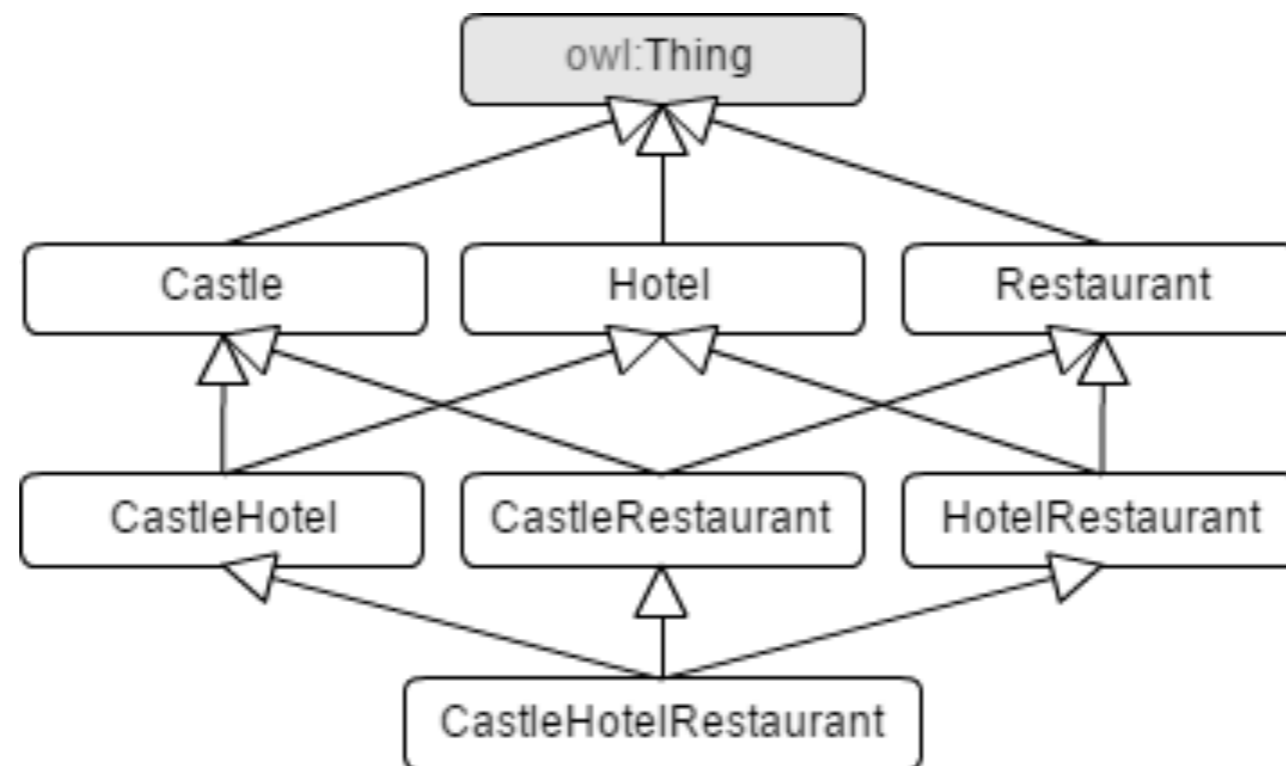


historic = castle

tourism = hotel

amenity = restaurant

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



# OF4OSM : une nouvelle base de connaissances OSM



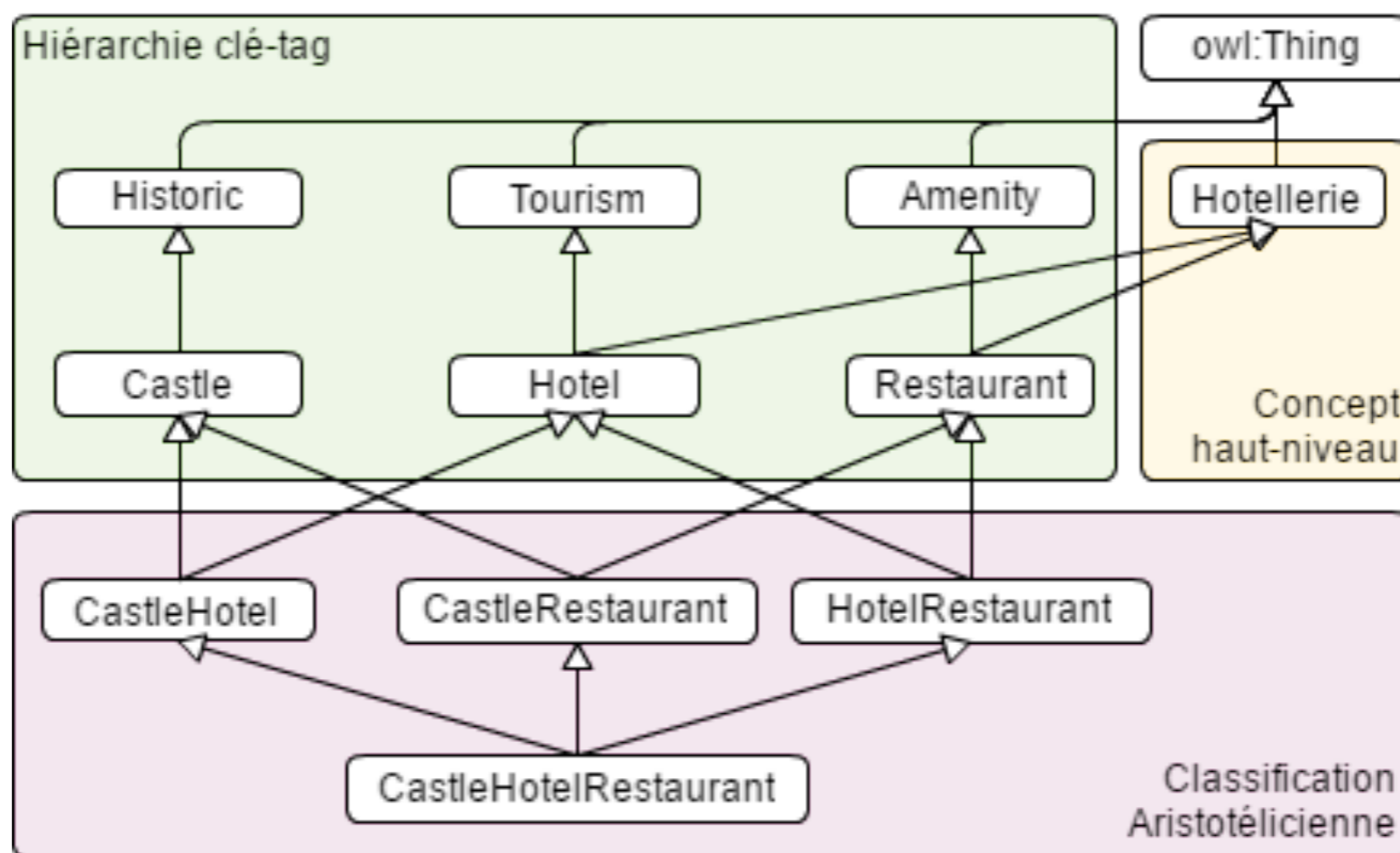
historic = castle

tourism = hotel

amenity = restaurant

Le Château de la Commanderie à Eybens en Isère

<http://www.openstreetmap.org/way/66383517>





# Plan

## Contexte et problématique

### I. Méthodes de structuration des tags OSM existantes

- ▶ Wiki OSM
- ▶ Linked Geo Data (LGD)
- ▶ OSMonto
- ▶ Game With A Purpose (GWAP)
- ▶ OSM Semantic Network (OSN)

### II. OF4OSM : nouvelle base de connaissances OSM

- ▶ Représentation d'un tag spécifique à OSM
- ▶ Typologie des tags OSM
- ▶ Génération de OF4OSM

## Bilan et perspectives

## Bilan

- ▶ Conception d'un **modèle de représentation des tags OSM**
- ▶ Conception d'une **typologie des tags OSM**
- ▶ Conception d'un **modèle de représentation des connaissances OSM** en 3 volets :
  - ▶ Hiérarchie clé-tag
  - ▶ Concepts haut niveau
  - ▶ Combinaisons de tags
- ▶ Développement d'**outils pour la génération d'une base de connaissances OSM** sous la forme d'une ontologie (OF4OSM)

## Perspectives

- ▶ Intégration d'OF4OSM dans un **greffon pour éditeur OSM** tel que OSMantic pour JOSM
  - ▶ Contrôle de la qualité des données lors de l'acquisition (suggestions de tags similaires, de combinaisons de tags, etc.)
- ▶ Conception d'un **outil pour l'exploration assistée** des données OSM basé sur OF4OSM
  - ▶ Aide à l'extraction de données (recherche de tags similaires, de tags co-occurents, intégration du contexte de l'utilisateur, etc.)
- ▶ Tests utilisateurs

- ▶ Auer, S. S., Lehmann, J., & 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) (Vol. 5823 LNCS, pp. 731–746). [https://doi.org/10.1007/978-3-642-04930-9\\_46](https://doi.org/10.1007/978-3-642-04930-9_46)
- ▶ Baglatzi, A., Kokla, M., & Kavouras, M. (2012). Semantifying OpenStreetMap. In The 11th International Semantic Web Conference (pp. 39–48).
- ▶ Ballatore, A., Bertolotto, M., & Wilson, D. C. (2013). Grounding linked open data in WordNet: The case of the OSM semantic network. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), 7820 LNCS, 1–15. [https://doi.org/10.1007/978-3-642-37087-8\\_1](https://doi.org/10.1007/978-3-642-37087-8_1)
- ▶ Bruns, A. (2006). Towards produsage: Futures for User-Led Content Production. In F. Sudweeks, H. Hrachovec, & C. Ess (Eds.), Cultural Attitudes towards Communication and Technology 2006 (pp. 275–284). Perth: Murdoch University.
- ▶ Codescu, M., Horsinka, G., Kutz, O., Mossakowski, T., & Rau, R. (2011). Osmonto-an ontology of openstreetmap tags. State of the Map. Retrieved from <http://www.informatik.uni-bremen.de/~okutz/osmonto.pdf>

# OF4OSM

## Une nouvelle base de connaissances extraite des tags OpenStreetMap

---

thèse de doctorat présentée par  
**Anthony Hombiat**

sous la direction de  
**M. Jérôme Gensel et Mme Marlène Villanova-Oliver**

**Équipe STeamer**  
**Laboratoire d'Informatique de Grenoble**

16 janvier 2017