



**HAL**  
open science

## Création d'un référentiel géo-historique d'adresses à partir de sources multiples

Charly Bernard, Nathalie Abadie, Julien Perret, Bertrand Duménieu

### ► To cite this version:

Charly Bernard, Nathalie Abadie, Julien Perret, Bertrand Duménieu. Création d'un référentiel géo-historique d'adresses à partir de sources multiples. GAST - Gestion et l'Analyse de données Spatiales et Temporelles, Jan 2024, Dijon, France. hal-04490732

**HAL Id: hal-04490732**

**<https://hal.science/hal-04490732>**

Submitted on 5 Mar 2024

**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.



Distributed under a Creative Commons Attribution - NonCommercial - ShareAlike 4.0 International License

# Création d'un référentiel géo-historique d'adresses à partir de sources multiples

Charly Bernard\*, Nathalie Abadie\*  
Julien Perret\*,\*\* Bertrand Duméniou\*\*

\*LASTIG, Université Gustave Eiffel, IGN/ENSG  
{charly.bernard ; nathalie-f.abadie ; julien.perret}@ign.fr,  
<https://www.umr-lastig.fr/>  
\*\*CRH, EHESS  
bertrand.dumenieu@ehess.fr

**Résumé.** Énoncés structurés désignant un lieu en s'appuyant sur son contexte géographique, les adresses servent aux individus à localiser une destination précise, un immeuble dans une ville par exemple, avec le moins d'ambiguïté possible. Ces références spatiales indirectes encodent des relations spatiales et hiérarchiques entre un nombre d'entités géographiques parfois important. Prises en masse, elles forment une description d'un territoire à différentes échelles. Lorsqu'il s'agit d'adresses anciennes, elles peuvent aider à reconstituer très finement les structures disparues d'un espace et ont un potentiel particulièrement puissant pour l'étude des dynamiques urbaines sur le temps long. La masse et l'hétérogénéité de ces objets rendent leur exploitation informatisée difficile malgré leur profusion dans les archives numérisées et dans les sources de données structurées sur le Web. Dans cet article, nous présentons une ontologie pour modéliser les adresses dans leur diversité et pour représenter l'évolution temporelle des entités géographiques qui les composent. Nous expliquons aussi comment et avec quelles données nous peuplons l'ontologie créée.

## 1 Introduction

La disponibilité croissante des sources d'archives numérisées et les progrès récents des approches d'extraction d'informations dans des textes ou des images, permis par l'essor des réseaux de neurones profonds, offrent aujourd'hui la possibilité d'exploiter ces sources pour produire des données structurées sur les divers aspects du passé qu'elles décrivent. Parmi les informations que l'on peut trouver dans les sources d'archives, les adresses constituent un corpus assez important à la fois par son volume et son intérêt applicatif. Premièrement, elles ont une place centrale dans divers types de documents iconographiques comme les plans de villes ou textuels comme les annuaires des habitants et des commerces, les actes d'état civil, les documents notariés, etc. Elles sont donc disponibles en grande quantité avec un taux de recouvrement potentiel entre sources relativement important. Deuxièmement, elles reflètent l'évolution des villes dont elles ont accompagné les changements morphologiques et administratifs

au cours du temps. Par exemple, à Paris, différents systèmes de numérotation des immeubles se sont succédé depuis l'Ancien Régime. Enfin, l'intérêt majeur des adresses réside dans leur double rôle de références spatiales directes, lorsqu'elles sont extraites de cartes et décrites par des coordonnées, et indirectes, quand elles sont représentées sous forme textuelle. En effet, disposer de cette double représentation permet de géocoder les sources d'archives textuelles et de localiser les informations qu'elles renferment à la surface de la Terre.

Afin de structurer et de valoriser les informations relatives aux adresses disponibles sous formes hétérogènes et fragmentaires dans les sources d'archives, nous proposons de construire une base de connaissances géohistorique sur les adresses de Paris. Le choix a été fait de se restreindre à la capitale française, car de nombreuses sources y sont disponibles, sur une période temporelle conséquente et représentent les adresses sous diverses formes. Cette base devra permettre de structurer les connaissances disponibles sur les adresses, de représenter leur évolution temporelle et les sources qui les mentionnent. Cet article présente l'ontologie développée pour représenter les deux premiers aspects des adresses historiques extraites de sources d'archives. Nous l'avons développée en suivant la méthodologie appelée *Simplified Agile Methodology for Ontology Development* dite SAMOD [Per17], qui consiste à séparer un problème de modélisation complexe en sous-problèmes, appelés modelets, plus simples à traiter, valider et tester. Un modelet commence avec un argumentaire en langage naturel qui décrit le sous-problème à traiter, avec un glossaire explicitant les notions et termes importants en jeu. Chaque modelet s'accompagne d'un ensemble de questions informelles de compétence, également en langage naturel, qui représentent les questions auxquelles la base de connaissances doit permettre de répondre. À chaque question est enfin associé un ensemble d'exemple de réponses attendues, qui servent à valider le modelet une fois implémenté. La modélisation de phénomènes géo-historiques est complexe et mobilise des thèmes différents de l'ingénierie des connaissances, la méthode SAMOD permet de les identifier et de les séparer pour faciliter la conception. En outre, cela permet d'opérer en cycles rapides avec une mise à l'épreuve régulière de l'ontologie en cours de construction.

Cet article est organisé de la façon suivante : nous présentons tout d'abord un état de l'art sur la représentation de connaissances sur les adresses et sur l'évolution temporelle des entités géographiques. Puis, nous présentons les deux modelets développés pour représenter des adresses de tous types et leurs évolutions au cours du temps. Enfin, nous peuplons ces modelets à l'aide de données réelles issues de diverses sources pour nous assurer de leur généralité et de leur capacité effective à répondre aux questions de compétences initialement définies.

## 2 État de l'art

### 2.1 Qu'est-ce qu'une adresse ?

Une adresse est avant tout une référence spatiale indirecte, un énoncé structuré qui désigne un lieu sans ambiguïté [CCD11]. Si la forme et la nature de ce qu'elle désigne varient selon les époques et les lieux, [Mor80] identifie deux catégories. L'adresse administrative, qui définit seulement la position d'un lieu, se distingue de l'adresse cheminement qui décrit un parcours dans l'espace. « 59 rue de Rivoli, 75001 Paris » et « maison de la Belle-Jardinière, quai aux Fleurs, coin de la rue de la Cité » sont respectivement des exemples de ces deux catégories d'adresses. Une adresse est en tout cas une entité composite, faite de références spatiales agen-

cées en hiérarchies et dont les relations spatiales sont éventuellement explicitées [CCD11]. Ces composantes sont des points de repères spatiaux, possiblement informels, connus et partagés par les "adressés" [Dum21]. Dans cet article, une adresse est donc un énoncé structuré d'un cheminement à l'intérieur d'une hiérarchie spatiale, non ambigu au sein de cette hiérarchie, composé d'une suite ordonnée de repères spatiaux dont la conceptualisation et la désignation sont connues et partagées.

## 2.2 Systèmes d'adressages à Paris depuis le XVIII<sup>e</sup> siècle

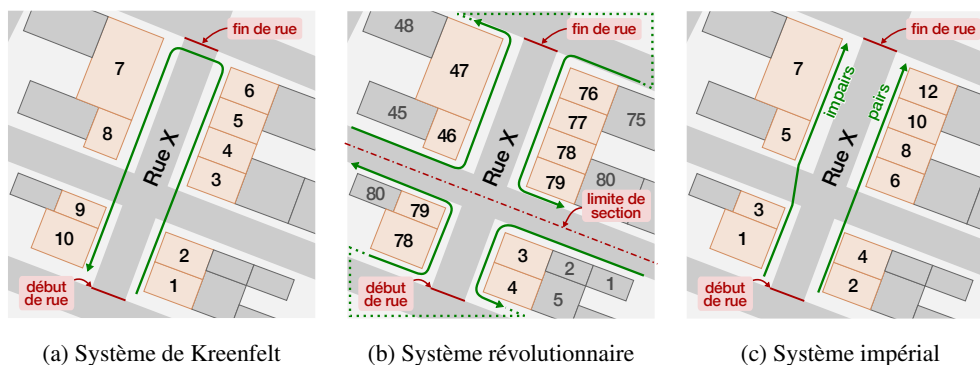


FIG. 1 – Les trois systèmes de numérotation des habitations utilisés à Paris. Les flèches vertes représentent le sens de numérotation.

Comme dans d'autres capitales européennes, l'adressage des maisons de Paris par des numéros apparaît à la fin du XVIII<sup>e</sup> siècle. Jusque-là, les bâtiments étaient désignés par un nom, souvent une enseigne, parfois avec des précisions pour le désambiguïser [Pro66], le tout complété de positionnements relatifs, comme la « *Maison du Pied de Griffon [...] au coin oriental de la rue du Chantre, [...] front sur la rue de Beauvais* »<sup>1</sup>. Trois systèmes de numérotation se succèdent ensuite. Un premier, d'initiative privée, est tenté par l'éditeur de l'*Almanach de Paris*<sup>2</sup>, Marin Kreenfelt, en 1779 [Dep16]. Les portes des bâtiments d'une même rue sont numérotées en suivant un parcours en « fer à cheval » (fig. 1a). Sans doute peu usitées dans les faits, ces adresses sont données dans l'*Almanach de Paris*. L'adressage dit *sectionnaire* apparaît durant la Révolution et ne lui survira pas longtemps. Nommé d'après la subdivision de Paris en sections, il s'appuie sur ce découpage. Créée dans un but surtout fiscal, cette numérotation n'a jamais été entièrement reconstituée, car elle s'appuyait sur une règle générale inégalement appliquée et sur de nombreuses adaptations locales non documentées [Fle95, Waq18]. Le numérotage débutait à une extrémité d'une section puis suivait un parcours le long des îlots urbains la composant, avec une logique propre à chacune (fig. 1b). Le troisième mode de numérotage, encore en vigueur aujourd'hui, est créé sous le Premier Empire. Il consiste à

1. Adolphe Bert, Henri Legrand, Lazare-Maurice Tisserand, Théodore Vacquer, Camille Platon. *Topographie historique du vieux Paris. Région du Louvre et des Tuileries*, volume 1. Paris, imprimerie impériale édition, 1866.

2. Marin Kreenfelt. *Almanach de Paris*, 1<sup>re</sup> partie contenant la demeure, les noms et qualités des personnes de condition, etc. 2<sup>e</sup> partie contenant les noms et demeures des principaux artistes, marchands, fabricants, etc. Pour l'année 1789. 1789.

numéroter les unités foncières (parcelle, maisons) de part et d'autre de chaque voie en ordre croissant, numéros impairs à gauche, numéros pairs à droite. Le point de départ de la rue et donc de sa numérotation est, en principe, l'extrémité la plus proche de la Seine (fig. 1c).

### 2.3 Modéliser les adresses

Si la littérature s'accorde sur la structure générale d'une adresse comme étant un ensemble organisé de points de repère reliés entre eux par des relations spatiales, les adresses postales concentrent l'attention du fait de leur importance administrative. Les propositions de représentation formelles existantes s'occupent essentiellement de ce type d'adresse, notamment dans un but d'harmonisation et de standardisation au sein d'un pays ou d'une région [CCD11, CC07, Zan08, BDS00].

Dans le domaine de la géomatique, la norme ISO 19133 [Iso05] porte uniquement sur des adresses postales. Les modèles à visée générique restent peu adaptés à des modes d'adressages plus hétérogènes. Par exemple, le modèle relationnel proposé par [DF07] décompose les adresses en repères géographiques reliés par des relations spatiales, mais avec une structure hiérarchique systématique : bâtiment, municipalité, région. C'est également le cas de l'ontologie *locn*<sup>3</sup>, dont la classe `locn:Address` est une accumulation de `Literal` et non de ressources. Les types de repères *y* sont limités (rue, unité administrative...) et les relations spatiales ne sont pas indiquées.

Ces propositions, prévues pour le cas des adresses postales contemporaines, ne sont pas immédiatement adaptables aux adresses anciennes de Paris [CDA<sup>+</sup>18]. Par exemple, le libellé d'adresse « *situé sur le blv. de Clichy, dans la partie comprise entre la place Blanche et la rue Fontaine* » ne peut pas être représenté par ces modèles : aucune municipalité n'est indiquée, il y a strictement plus d'un repère de type Voie et il y a une relation spatiale définie en langage naturel par le terme *entre*. Enfin, aucun de ces modèles n'intègre le fait que les adresses évoluent au cours du temps.

### 2.4 Représenter l'évolution temporelle des adresses

La représentation de dynamiques spatiales a fait l'objet de plusieurs recherches [Hal12, CTP98, FYZW10, DM11]. Ces travaux ont en commun une question prégnante : qu'est-ce qui définit l'identité d'une entité spatio-temporelle ? Par exemple, à partir de quand considère-t-on qu'une entité est créée ou a disparu ? [Hal12] propose de dissocier l'identité d'une entité de sa réalité physique. Ainsi, une rue n'est pas créée dès lors qu'elle existe sur le terrain, mais dès sa planification. L'évolution temporelle qu'il propose s'appuie sur le modèle de l'*Identity Based Change* [HE00]. Le principe repose sur l'alternance de phases (ou états spatio-temporels) d'une entité durant sa vie. Les changements sont définis selon la transition entre deux phases. [HE00] mentionne trois états que peut avoir une entité géographique : l'existence, la non-existence simple et la non-existence avec une histoire ; [Hal12] en propose cinq. Avec ces approches, on peut supposer que les entités ont des durées de vie infinies qui possèdent divers états d'existence encadrés par des états de non-existence.

Il existe différentes approches pour former un graphe géohistorique. Cependant, les principales sont limitées à des données homogènes à temporalités fixées [Arm88, Dum15, Cos16,

---

3. <http://www.w3.org/ns/locn>

BVOGD18]. Elles reposent sur la détection des versions d’une entité géographique au sein des sources indépendantes qui décrivent chacune un état spatio-temporel. S’il existe des différences entre les versions successives, des changements en sont déduits. C’est le cas de l’ontologie TSN-Change [BVOGD18] qui permet de tracer l’évolution d’une entité géographique (apparition, disparition, changement de nom, etc.) au sein des différentes versions de la NUTS<sup>4</sup>. Cette nomenclature, publiée chaque année, est un découpage multiniveau du territoire de l’Espace économique européen. L’originalité de l’approche est que les changements peuvent inclure plusieurs entités (fusion, scission. . .) en les divisant en changements élémentaires qu’on applique au niveau des entités. [BCH<sup>+</sup>22] propose l’ontologie HHT (Historical Hierarchical Territory) qui applique une approche similaire aux territoires de l’Ancien Régime.

Pour représenter formellement la dimension temporelle de données liées, le W3C propose l’ontologie OWL-time qui permet d’exprimer des faits liés aux ressources avec une approche temporelle. Implémentant l’algèbre temporelle de Allen, elle permet d’associer un événement à un instant ou un intervalle de temps et d’ordonner des faits temporellement [PH06].

### 3 Proposition

#### 3.1 Adresses

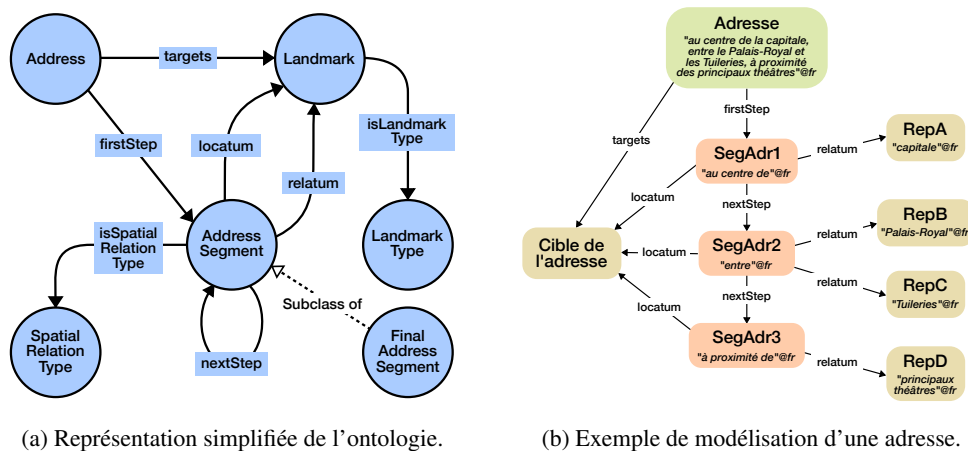


FIG. 2 – Modélisation des adresses.

Dans ce modèle<sup>5</sup>, une adresse est composée de références à des repères spatiaux, liés par des relations spatiales et qui ensemble pointent vers un lieu cible auquel on peut associer des coordonnées géographiques. Ainsi, nous avons retenu les questions de compétences suivantes : (1) quelles sont les adresses répertoriées le long d’une rue donnée ? (2) quelles sont les coordonnées

4. Nomenclature des unités territoriales statistiques, voir <https://ec.europa.eu/eurostat/fr/web/nuts/background>

5. <https://github.com/charlybernard/phd-ontologie>

## Création d'un référentiel géo-historique d'adresses à partir de sources multiples

données correspondant à la cible d'une adresse donnée ? (3) quelles sont les adresses localisées dans une zone donnée ? L'ontologie, présentée en figure 2a, comporte trois concepts :

- l'adresse (*Address*), qui correspond au libellé décrivant une localisation ;
- le segment d'adresse (*AddressSegment*) décrivant une partie du cheminement d'une adresse en reliant des repères via une relation spatiale (*SpatialRelationType*) ;
- le repère (*Landmark*) qui décrit une entité géographique. *LandmarkType* définit sa nature (par exemple : unité administrative, voie, numéro d'immeuble, bâtiment...).

Le lieu désigné par l'adresse est dénommé ici *Landmark* et la suite des repères géographiques qui composent l'adresse sont des ressources de type *AddressSegment*. Une ressource de type *Address* est liée via la propriété *firstStep* à une première ressource de type *AddressSegment* qui est elle-même liée via la propriété *nextStep* à une ressource du même type et ainsi de suite. Pour décrire la fin de cette suite, la dernière ressource est de type *FinalAddressSegment* (segment d'adresse final), sous-classe de *AddressSegment*. Une ressource de type *AddressSegment* décrit une étape du cheminement qui lie un repère, qu'on appelle *locatum*, à un ou plusieurs repères qu'on nomme *relatum* [TK11]. Généralement, le *locatum* est la cible de l'adresse, car chaque étape du cheminement a pour objectif de donner des précisions sur la localisation de la cible. Prenons l'exemple présenté par la figure 2b. L'adresse cible un repère inconnu et a un libellé qu'on peut découper en trois parties : « au centre de la capitale », « entre le Palais-Royal et les Tuileries » et « à proximité des principaux théâtres ». Il y a trois segments et quatre références à des repères géographiques.

### 3.2 Évolution temporelle

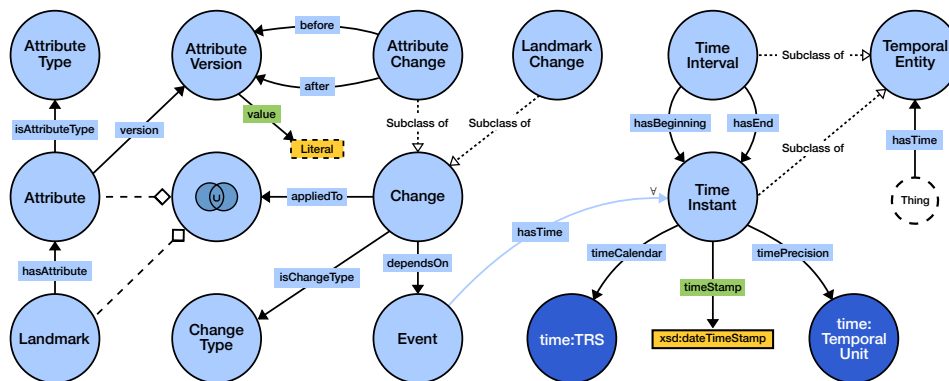


FIG. 3 – Représentation de l'ontologie de l'évolution temporelle des repères géographiques<sup>6</sup>.

Pour ce modèle, nous avons identifié les questions de compétences suivantes : (1) quelles sont les voies existant à un instant donné ? (2) sur quel(s) intervalle(s) temporel(s) est valable une adresse de dénomination donnée ? (3) quel est l'historique d'un repère ? Autrement dit,

6. L'élément lié à la classe *Change* par le prédicat *appliedTo* décrit l'union de *Attribute* et de *Landmark*.

quels sont les événements qui lui sont associés ? (4) quels sont les états et les événements qui manquent dans l'historique d'une adresse ? Deux besoins principaux se dégagent de ces questions de compétences : pouvoir retrouver à chaque instant l'état des repères géographiques et représenter les événements ayant eu un effet sur les repères géographiques et leurs propriétés. Pour ce deuxième aspect, on s'inspire des travaux de [BVOGD18] en divisant les événements en changements élémentaires qui explicitent la succession de deux versions. En plus de s'appliquer aux repères, nous avons fait le choix de lier les changements aux attributs des repères. Ainsi, il est possible de déterminer l'évolution d'un attribut en particulier. Pour cela, une propriété liée à un repère est définie via une ressource de type `Attribute` à qui on associe des versions. Dès qu'un changement est appliqué au repère ou à un de ses attributs, on l'indique via une ressource de type `Change` qui décrit à quelle ressource s'applique le changement. [BVOGD18] étudie l'évolution d'entités au sein de *snapshots*, donc les changements décrivent l'évolution entre deux versions successives d'une entité même si aucune différence n'existe. L'approche est discrète, or nous voulons que chaque entité ait un temps de validité qui lui soit propre pour décrire l'état du territoire à chaque instant, ce qui est possible grâce à notre modélisation. La figure 3 présente les quatre concepts principaux de l'ontologie :

1. le changement (`Change`) décrit une opération élémentaire d'un événement. S'il s'applique à un repère géographique, le changement est un `LandmarkChange` et s'il s'applique à un attribut (comme le nom d'une rue), il est un `AttributeChange`. `ChangeType` définit son type (apparition ou disparition d'un repère par exemple) ;
2. l'attribut (`Attribute`) dont le type est donné par `AttributeType` ;
3. la version d'attribut (`AttributeVersion`).

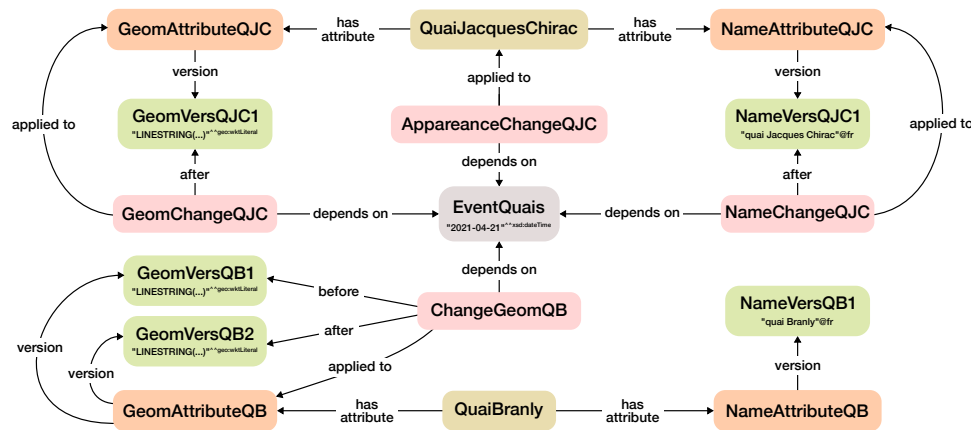


FIG. 4 – Exemple de modélisation de l'évolution conjointe de deux voies : création du *quai Jacques Chirac* par démembrement du *quai Branly*.

Avec cette modélisation, nous pouvons découper des modifications impliquant une ou plusieurs entités géographiques en changements élémentaires. Prenons l'exemple avec l'événement



ment « création du quai Jacques Chirac sur une portion du quai Branly à la date du 21 avril 2021 ». Nous pouvons le diviser en plusieurs changements comme représentés sur la figure 4 :

- apparition du quai Jacques Chirac (`QuaiJacquesChirac`);
- apparition d'une nouvelle version pour `NameAttributeQJC`, attribut de type nom, la version précédente n'existe pas ;
- apparition d'une nouvelle version pour `GeomAttributeQJC`, attribut de type géométrie, la version précédente n'existe pas non plus ;
- apparition d'une nouvelle version pour `GeomAttributeQB`, attribut de type géométrie du quai Branly, la version précédente étant l'emprise spatiale du quai allant de la place de la Résistance au pont d'Iéna.

## 4 Mise en œuvre et évaluation

La cohérence et le respect des bonnes pratiques de conception des ontologies proposées ont été vérifiés à l'aide du raisonneur `Hermit` intégré à `Protégé`<sup>7</sup> [Mus15] ainsi que de l'outil `OOPS!`<sup>8</sup> [PVGPSF14].

### 4.1 Peuplement et validation de l'ontologie des adresses

L'ontologie des adresses est également évaluée en la peuplant avec trois sources de données d'adresses sur Paris ; deux contemporaines : `OpenStreetMap` (OSM) et la Base Adresse Nationale française (BAN), et une historique : les adresses extraites des annuaires du commerce de Paris du XIX<sup>e</sup> siècle [ACCD22]. Le processus d'intégration de ces sources comprend trois étapes : (1) identifier les repères et les relations spatiales composant les adresses, (2) structurer chaque adresse selon le modèle proposé, (3) lier les adresses désignant les mêmes cibles.

num	rep	nom_voie	cp	nom_com	lat	lon
5		Boulevard Saint-Martin	75003	Paris 3e	2.361407	48.867933
134		Boulevard Raspail	75006	Paris 6e	2.309998	48.86011
4		Rue Regnard	75006	Paris 6e	2.338073	48.849885
48	bis	Rue de Rivoli	75004	Paris 4e	2.35448	48.85690

TAB. 1 – Extrait d'un fichier de la BAN pour la ville de Paris<sup>9</sup>.

La première étape est triviale pour les données OSM et BAN, composées d'adresses postales dont les éléments sont déjà séparés. La structuration des adresses historiques des annuaires du commerce de Paris, plus organiques (figure 2b), est difficile et des outils de *par-sing* sur étagère comme `LibPostal`<sup>10</sup> échouent. Nous avons donc structuré manuellement ces

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

8. <https://oops.linkeddata.es/>

9. Les positions des adresses (fournies par `lat` et `lon`) dans la BAN n'a pas de règle unique définie. Il est indiqué pour chaque adresse la manière dont est fixée cette position (par rapport au point de livraison postal, à la porte d'entrée, au centre de la parcelle, etc).

10. <https://github.com/openvenues/libpostal>

adresses. Les données de chaque source sont ensuite formatées en CSV ou en JSON, éventuellement après une restructuration légère. Par exemple, les données de la BAN (table 1) sont adaptées pour créer un champ `house_number` qui concatène les champs `num` et `rep`.

Un mapping pour chaque source décrit la manière dont les adresses doivent être converties en RDF, conformément à l'ontologie créée : il précise les repères et les relations spatiales entrant en jeu. Nous utilisons le logiciel Ontotext Refine<sup>11</sup> pour construire ces adresses. Le mapping d'un fichier de la BAN permet de créer neuf ressources : une `Address`, quatre `AddressSegment`, quatre `Landmark` (`HouseNumber`, `Thoroughfare`, `PostalCode` et `City`). Nous obtenons ainsi un graphe d'adresses, mais, les adresses étant créées de manière indépendante, les repères entrant en jeu y sont dupliqués : pour les adresses situées le long d'une même rue, on crée autant de ressources différentes pour cette rue que d'adresses. La dernière étape vise donc à éliminer les doublons à l'aide de requêtes SPARQL : les repères de même type, ayant des noms proches et des relations spatiales similaires, sont fusionnés.

Une fois peuplée, nous avons vérifié que l'ontologie permet de répondre aux questions de compétence prédéfinies, transcrites en requêtes SPARQL. Ainsi, la requête pour sélectionner l'ensemble des adresses ciblant un lieu situé le long d'une voie donnée (ici : la rue du Dahomey) renvoie 16 adresses et leurs coordonnées, cartographiées en figure 5.

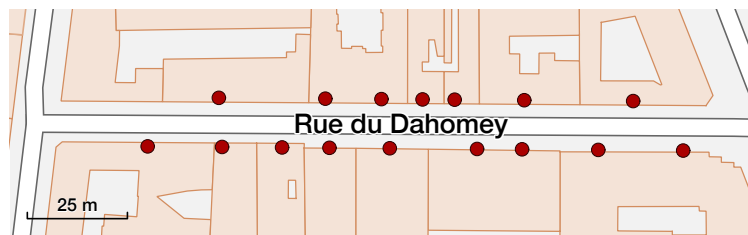


FIG. 5 – Localisation des adresses dont la cible est située le long de la rue du Dahomey.

## 4.2 Peuplement et validation de l'ontologie des évolutions temporelles

Pour peupler l'ontologie sur l'évolution temporelle des repères géographiques, nous avons utilisé des données de Wikidata sur les voies de Paris. Elles contiennent des informations comme la date de création, une liste des noms officiels avec des dates de début et fin de validité. À partir de ces données, nous créons des repères de type `Thoroughfare`, des attributs de noms avec leurs différentes versions et leurs changements. Wikidata décrit les états des voies. Donc, pour chacun d'entre eux, nous créons deux changements indiquant respectivement le début et la fin de validité de l'état. Cela implique que tous les changements sont indépendants et peuvent être dupliqués. Il faut donc procéder à des fusions via des requêtes SPARQL.

Ce jeu de données se limitant à l'historique des voies, seules certaines questions vues en section 3.2 ont été transcrites en requêtes SPARQL. L'historique d'une entité géographique peut être obtenu soit en requêtant tous les changements opérés sur l'entité, soit en construisant des versions en fonction des changements existants, comme illustré en figure 6. Ici, les états sont déduits des événements et peuvent différer selon les événements qu'on y inclut ou pas.

11. <https://www.ontotext.com/products/ontotext-refine/>

## Création d'un référentiel géo-historique d'adresses à partir de sources multiples

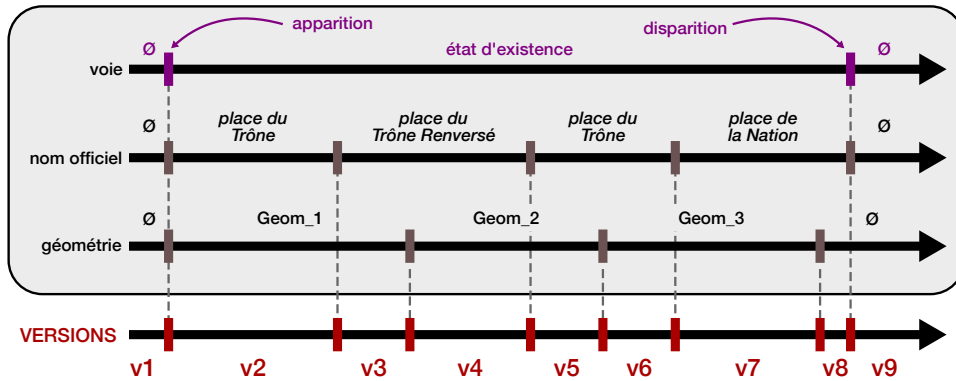


FIG. 6 – Reconstruction des états de la place de la Nation à partir des changements appliqués.

## 5 Discussion

Dans cet article, nous avons présenté deux ontologies : une pour modéliser les adresses dans leur diversité et une pour représenter l'évolution temporelle des entités géographiques qui les composent. Elle permet la construction cumulative d'un graphe géohistorique à partir de données fragmentaires. Il reste toutefois quelques limites à cette modélisation dont la résolution devra faire l'objet de travaux futurs. Premièrement, nous avons associé aux événements des valeurs temporelles qui sont considérées comme connues et précises. Or, l'imprécision fait partie intégrante des données historiques, et il faut donc prendre en compte cet aspect. Deuxièmement, l'ontologie ne permet pas pour le moment de traiter l'évolution temporelle des relations (notamment spatiales) existant entre les repères géographiques.

En suivant la méthode SAMOD, il reste un modèle à traiter qui concerne les sources et les affirmations qui y sont contenues. Il permettra d'alimenter le graphe à partir de différents fragments d'informations de sources hétérogènes et aussi de gérer les conflits qui peuvent exister entre des sources potentiellement contradictoires.

Enfin, pour valider l'ontologie proposée, nous souhaitons la peupler plus largement, avec différents types de sources de données sur Paris : des sources cartographiques anciennes, dotées de géométries, de type *snapshot*, comme les atlas de Verniquet, de Jacoubet, etc. ; des sources structurées à différentes temporalités comme les *Dénominations caduques des voies de Paris*<sup>12</sup> ou la BAN<sup>13</sup> ; des sources non structurées, également à différentes temporalités, comme des annuaires anciens, ou les descriptions textuelles de l'histoire des rues fournies dans Wikipedia. Le graphe de connaissance ainsi peuplé devra permettre d'alimenter un outil de géocodage historique, capable de localiser des intitulés d'adresses dans l'espace parisien et dans leur période d'existence connue.

12. <https://opendata.paris.fr/>

13. <https://www.data.gouv.fr/fr/datasets/base-adresse-nationale/>

## Références

- [ACCD22] N. Abadie, E. Carlinet, J. Chazalon, and B. Duméniéu. A Benchmark of Named Entity Recognition Approaches in Historical Documents Application to 19th Century French Directories. In Seiichi Uchida, Elisa Barney, and Véronique Eglin, editors, *Document Analysis Systems*, Lecture Notes in Computer Science, pages 445–460, Cham, 2022. Springer International Publishing.
- [Arm88] Marc P Armstrong. Temporality in spatial databases. *GIS/LIS 88 Proceedings : Accessing the world*, pages 880–889, 1988.
- [BCH<sup>+</sup>22] Lucas Bourel, William Charles, Nathalie Jane Hernandez, Nathalie Aussenac-Gilles, Victor Gay, and Sébastien Poublanc. Graphes de connaissances pour représenter et analyser l'évolution des territoires en Histoire. In Nicolas Lasolle, Olivier Bruneau, and Jean Lieber, editors, *Journées humanités numériques et Web sémantique 2022*, pages 23–37, Nancy, France, June 2022. Zenodo. Backup Publisher : Laboratoire AHP-PReST : Archives Henri-Poincaré - Philosophie et Recherches sur les Sciences et les Technologies and LORIA : Laboratoire Lorrain de Recherche en Informatique et ses Applications.
- [BDS00] Vinayak Borkar, Kaustubh Deshmukh, and Sunita Sarawagi. Automatically Extracting Structure from Free Text Addresses. *IEEE Data Eng. Bull.*, 23 :27–32, January 2000.
- [BVOGD18] Camille Bernard, Marlène Villanova-Oliver, Jérôme Gensel, and Hy Dao. Ontologies pour représenter l'évolution des découpages territoriaux statistiques. *Revue Internationale de Géomatique*, 28(4) :409–437, December 2018.
- [CC07] Serena Coetzee and Antony K. Cooper. What is an address in South Africa? *South African Journal of Science*, 103(11-12) :449–458, December 2007. Publisher : Academy of Science of South Africa.
- [CCD11] Serena Coetzee, Antony K Cooper, and Jeofrey Ditsela. Towards good principles for the design of a national addressing scheme. In *25th International Cartographic Conference (ICC 2011)*. French Committee of Cartography Paris, France, 2011.
- [CDA<sup>+</sup>18] Rémi Cura, Bertrand Dumenieu, Nathalie Abadie, Benoit Costes, Julien Perret, and Maurizio Gribaudo. Historical Collaborative Geocoding. *ISPRS International Journal of Geo-Information*, 7(7) :262, July 2018. Number : 7 Publisher : Multidisciplinary Digital Publishing Institute.
- [Cos16] Benoît Costes. *Vers la construction d'un référentiel géographique ancien : un modèle de graphe agrégé pour intégrer, qualifier et analyser des réseaux géohistoriques*. PhD Thesis, Université Paris-Est, November 2016.
- [CTP98] Christophe Claramunt, Marius Thériault, and Christine Parent. A qualitative representation of evolving spatial entities in two-dimensional topological spaces. In *Innovations In GIS 5*, pages 128–142. CRC Press, March 1998.
- [Dep16] Thierry Depaulis. Vaugeois, tabletier, connu et inconnu (1re partie). *Le Vieux Papier*, 41(421) :106–113, July 2016.
- [DF07] Clodoveu A. Davis and Frederico T. Fonseca. Assessing the Certainty of Loca-

## Création d'un référentiel géo-historique d'adresses à partir de sources multiples

- tions Produced by an Address Geocoding System. GeoInformatica, 11(1) :103–129, March 2007.
- [DM11] Géraldine Del Mondo. Un modèle de graphe spatio-temporel pour représenter l'évolution d'entités géographiques. phdthesis, Université de Bretagne occidentale, Brest, October 2011.
- [Dum15] Bertrand Dumenieu. Un système d'information géographique pour le suivi d'objets historiques urbains à travers l'espace et le temps. PhD Thesis, École des Hautes Études en Sciences Sociales, December 2015.
- [Dum21] Gift Dumedah. Address points of landmarks and paratransit services as a credible reference database for geocoding. Transactions in GIS, 25(2) :1027–1048, 2021.
- [Fle95] M. Fleury. Concordance entre la numérotation sectionnaire et la numérotation de type actuel de 2466 maisons de Paris. Commission du Vieux, Rotonde de La Villette, Institut d'histoire de Paris, 1995.
- [FYZW10] Y. T. Fan, J. Y. Yang, D. H. Zhu, and K. L. Wei. A time-based integration method of spatio-temporal data at spatial database level. Mathematical and Computer Modelling, 51(11) :1286–1292, June 2010.
- [Hal12] Pierre Hallot. L'identité à travers l'espace et le temps. Vers une définition de l'identité et des relations spatio-temporelles entre objets géographiques. PhD Thesis, ULiège - Université de Liège, March 2012.
- [HE00] Kathleen Hornsby and Max J. Egenhofer. Identity-based change : a foundation for spatio-temporal knowledge representation. International Journal of Geographical Information Science, 14(3) :207–224, April 2000.
- [Iso05] ISO Iso. 19133 : Geographic information-Location-based services. Tracking and navigation. International Organization for Standardization. Technical Report, 2005.
- [Mor80] Takashi Morita. Cheminement et adresse. Les Annales de la Recherche Urbaine, 7(1) :27–61, 1980. Publisher : Persée - Portail des revues scientifiques en SHS.
- [Mus15] Mark A. Musen. The Protégé Project : A Look Back and a Look Forward. AI matters, 1(4) :4–12, June 2015.
- [Per17] Silvio Peroni. A Simplified Agile Methodology for Ontology Development. In Mauro Dragoni, María Poveda-Villalón, and Ernesto Jimenez-Ruiz, editors, OWL : Experiences and Directions – Reasoner Evaluation, Lecture Notes in Computer Science, pages 55–69, Cham, 2017. Springer International Publishing.
- [PH06] Feng Pan and Jerry R Hobbs. Time Ontology in OWL. W3C working draft, W3C, 2006.
- [Pro66] Jeanne Pronteau. Les Numérotages des maisons de Paris, du XVe siècle à nos jours ... Commission des travaux historiques. Préfecture de la Seine, Service des travaux historiques, 1966.
- [PVGPSF14] María Poveda-Villalón, Asunción Gómez-Pérez, and Mari Carmen Suárez-Figueroa. OOPS! (OntOlogy Pitfall Scanner!) : An On-line Tool for Ontology

Evaluation. International Journal on Semantic Web and Information Systems (IJSWIS), 10(2) :7–34, 2014. Publisher : IGI Global.

- [TK11] Thora Tenbrink and Werner Kuhn. A Model of Spatial Reference Frames in Language. In Max Egenhofer, Nicholas Giudice, Reinhard Moratz, and Michael Worboys, editors, Spatial Information Theory, pages 371–390, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg.
- [Waq18] Dominique Waquet. Almanachs, cadastre, terriers clés du décodage des numéros sectionnaires d’immeubles à Paris, 1791 - 1805. Annales historiques de la Révolution française, 392(2) :173–183, 2018. Place : Paris Publisher : Armand Colin.
- [Zan08] Paul A. Zandbergen. A comparison of address point, parcel and street geocoding techniques. Computers, Environment and Urban Systems, 32(3) :214–232, May 2008.

## Summary

Addresses are structured designations of places that are used by individuals to locate a precise destination - a building in a city, for example - with as little ambiguity as possible. These indirect spatial references encode spatial and hierarchical relationships between a large number of geographical entities, and taken in mass form a description of an area at different scales. In the case of old addresses, they can help to reconstruct the vanished structures of an area in great detail, and have a particularly powerful potential for the study of urban dynamics over the long term. The mass and heterogeneity of these objects make them difficult to use in computerised form, despite their abundance in digitised archives and structured data sources on the Web. In this article, we present an ontology for modelling addresses in all their diversity and for representing the temporal evolution of geographical entities. We also explain how and with what data we populate the ontology we have created.