

PeGazUs : une méthode de reconstitution de l'évolution des entités géographiques à partir de données hétérogènes et fragmentaires

Charly Bernard¹

Nathalie Abadie¹, Bertrand Duménieu², Julien Perret^{1,2}

¹ LASTIG, Université Gustave Eiffel, IGN-ENSG, Saint-Mandé, France

² CRH, EHESS-CNRS, Paris, France

Contexte et objectifs

- Adresses décrivent des lieux qui évoluent
- Géocodage d'adresses anciennes avec les données actuelles est limité
- Création d'un référentiel géo-historique d'adresses

Constant, ébéniste, St-Dominique, 100.
Constant, ferrailleur, Rocher, 3.
Constant, fab. de garnit. de parapluies, cour
de la Trinité, 18.
Constant, graveur, boul. d'Italie, 35.
Constant, jardinier, pl. du Trône, 6, et boul.
Mazas, 114.
Constant (Leuckens), menuisier, fab. de châs-
sis à clefs, Marais-St-Germain, 13.
Constant (Mme), robes, St-Honoré, 249.
Constant (F.), vins. N.-D.-des-Victoires, 13.

Extrait d'un annuaire commercial de Paris de 1857

Verrous

- Aucune base de données / source ne décrit complètement l'évolution des entités géographiques (repères)
- Peupler le référentiel **itérativement** avec des données **fragmentaires et hétérogènes**

GÉRARD (passage)**

XIII^e Arrondissement. Commence 21 bd Auguste-Blanqui ; finit 22 r. Gérard. Longueur 25 m ; largeur 3,75 m.

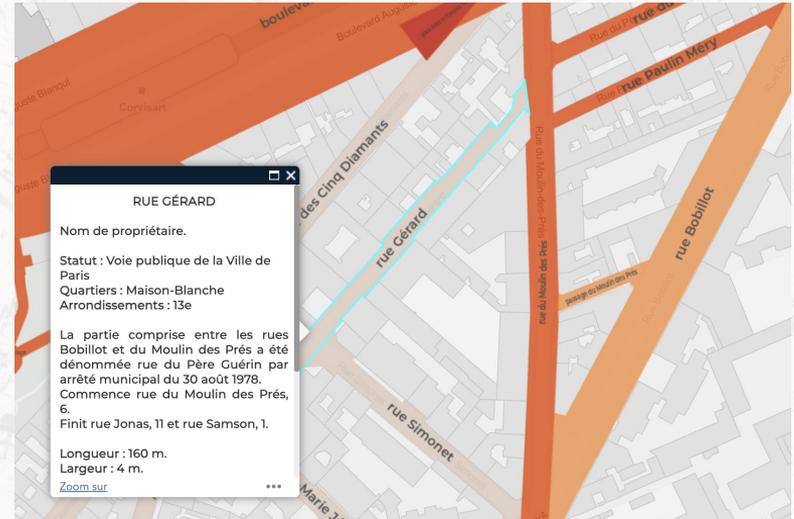
Cette voie privée s'est appelée *passage des Eaux* avant de recevoir, en 1877, son nom actuel, dû au voisinage de la rue désignée ci-après.

GÉRARD (rue)**

XIII^e Arrondissement. Commence 4 r. Bobillot ; finit 11 r. Jonas et 1 r. Samson. Longueur 307 m ; largeur 10 m. En partie, ancienne voie de la commune de Gentilly.

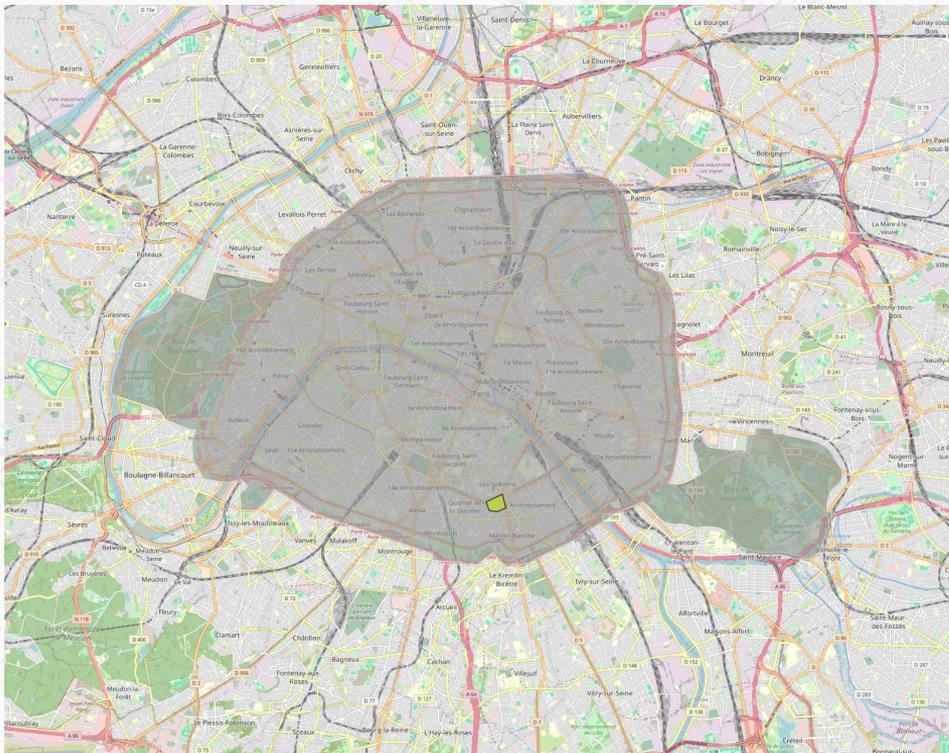
Cette rue, formée, en 1851, entre les rues Bobillot et du Moulin-des-Prés et, en 1857, entre cette dernière rue et la rue Jonas, porte le nom d'un propriétaire.

N° 36. — Ancien débouché de l'impasse Levé (voie supprimée).

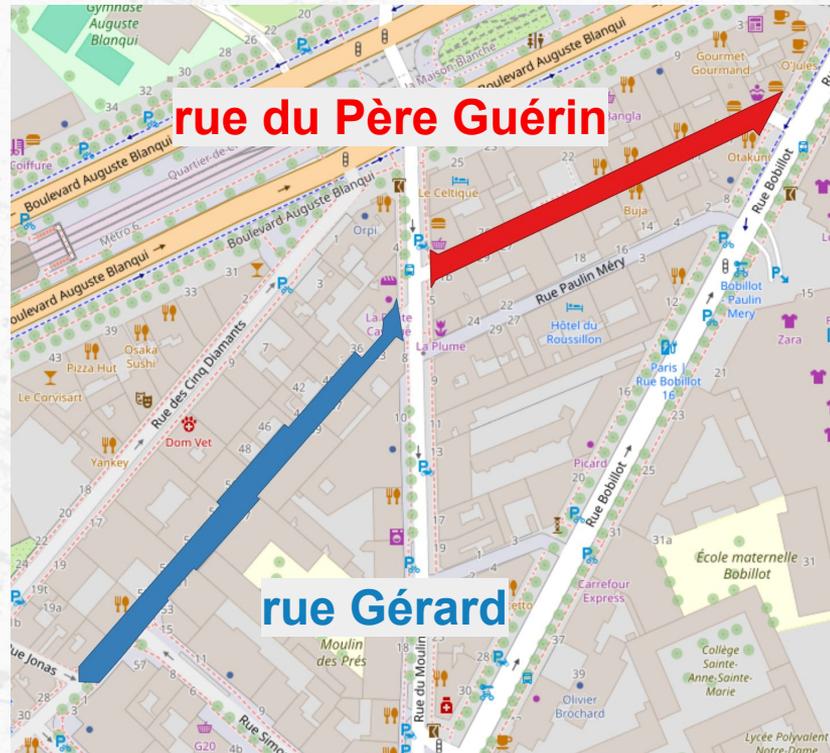


Cas d'application

Quartier de la Butte aux Cailles, Paris



Localisation de la Butte aux Cailles dans Paris



Localisation de deux rues de la Buttes aux Cailles

Cas d'application

Rue Gérard : voie de la Butte aux Cailles ayant subi plusieurs transformations



**Cadastré napoléonien
de Gentilly**

1845-1849



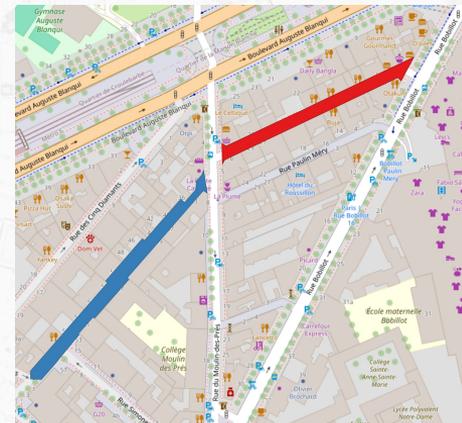
**Plan parcellaire
municipal**

1870-1872



Plan Andriveau

1887-1889



**Nomenclature de la
Ville de Paris**

2025

Feuille de route

- Créer une ontologie d'**adresses** et d'évolution spatio-temporelle d'**entités géographiques**
 - trois parties : **adresses, entités géographiques, sources**
 - **questions de compétences** : inférer l'état du territoire à chaque instant sans avoir de données décrivant cet instant
- **Population multi-source** du graphe à partir de données **fragmentaires et hétérogènes**
- **Évaluation** du graphe

Ontologie : Entités géographiques

Approches existantes

- Kauppinen et al. [1]: communes finlandaises (depuis 1865)
- Territorial Statistical Nomenclature (TSN) [2]: unités administratives et statistiques
- Hierarchical Historical Territory (HHT) [3]: subdivisions territoriales françaises de l'Ancien Régime

⇒ Concept **Change-Bridge** [1]: description de changements entre des versions d'entités géographiques

⇒ Données : **structurées, homogènes**

⇒ Versions **désambiguïsées**

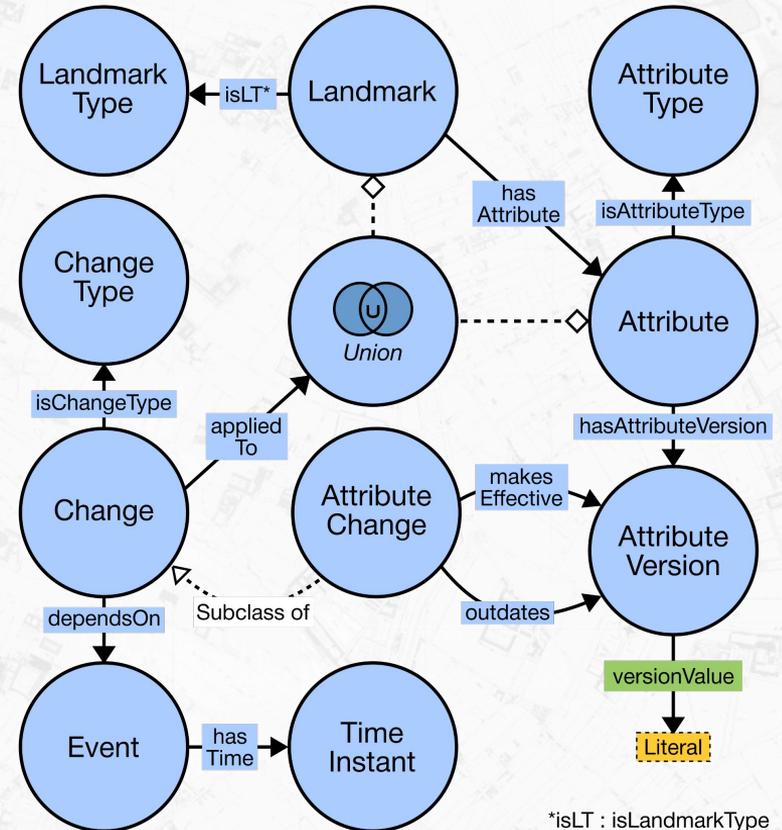
[2] Kauppinen et al. (2008). Creating and Using Geospatial Ontology Time Series in a Semantic Cultural Heritage Portal

[3] Bernard et al. (2018) Modeling changes in territorial partitions over time: Ontologies tsn and tsn-change

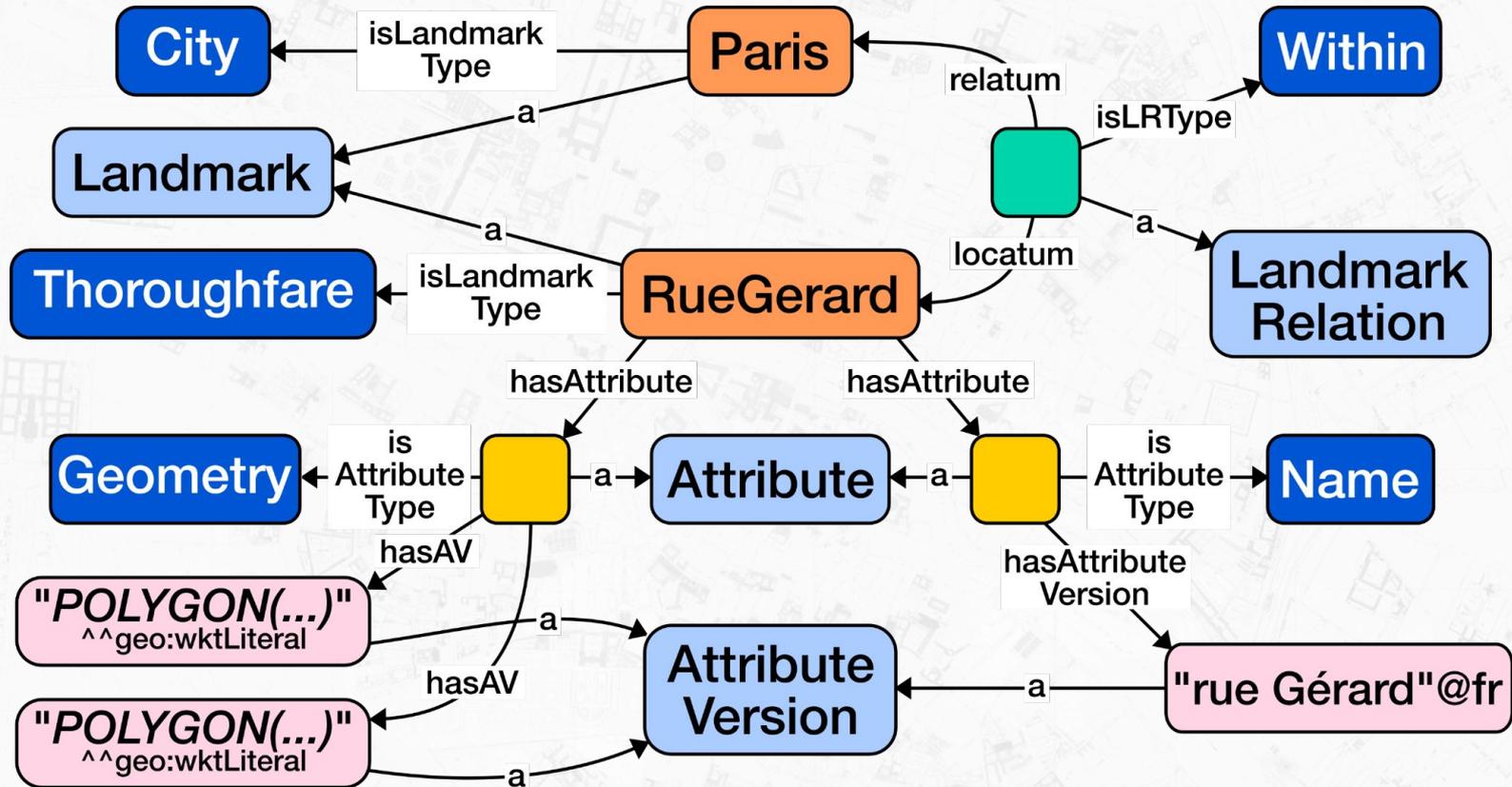
[4] Charles et al. (2023) : HHT: an approach for representing temporally-evolving historical territories

Ontologie : Entités géographiques

- Versionnement des attributs et non des entités
- Concept **Change-Bridge** [2, 3, 4] : description de changements entre des versions successives
 - **Event** : évolution du territoire
 - **Change** : évolution dans les données induite par un Event

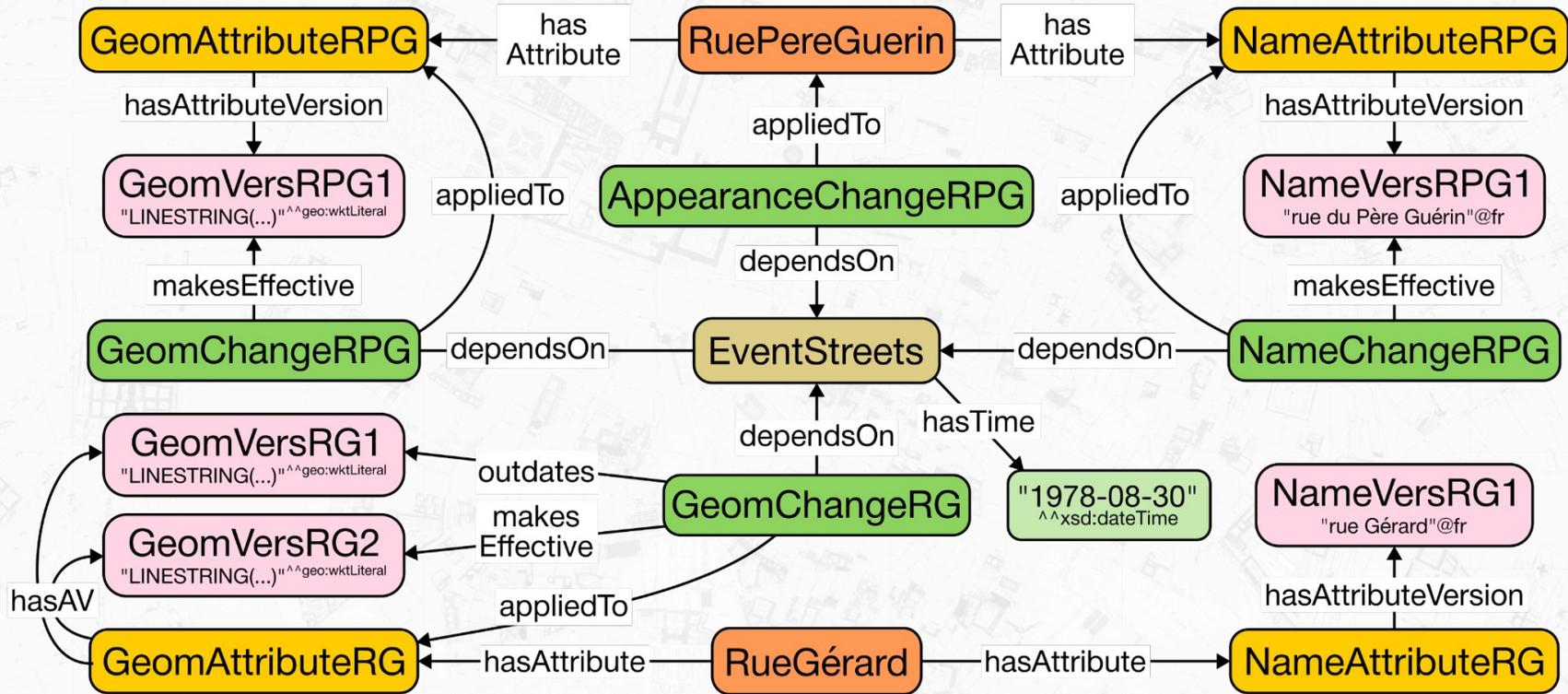


Ontologie : Entités géographiques



Représentation de la rue Gérard

Ontologie : Entités géographiques



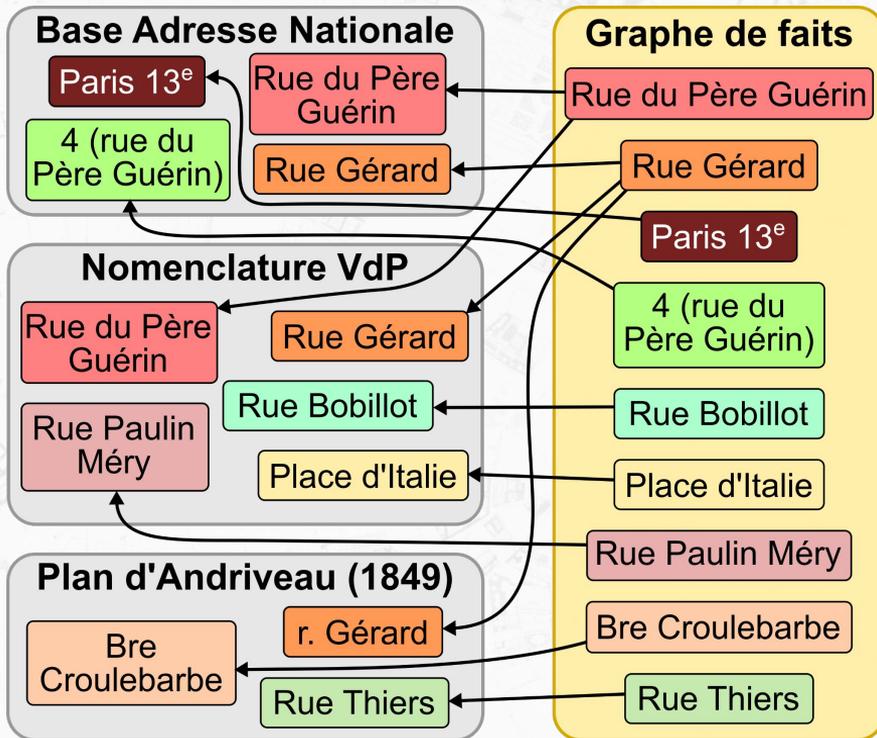
Représentation de l'événement "Par arrêté municipal du 30 août 1978, sa portion orientale, de la rue Bobillot à la rue du Moulin-des-Prés, prend le nom de rue du Père-Guérin."

Population

- Méthode de population proposée par [5]
- Limites :
 - peu adapté à l'intégration d'événements
 - problème d'inférences de changements
- Présentation d'une méthode plus robuste

[5] Charly Bernard, Solenn Tual, Nathalie Abadie, Bertrand Duméniou, Julien Perret, et al.. PeGazUs: A knowledge graph based approach to build urban perpetual gazetteers. International Conference on Knowledge Engineering and Knowledge Management (EKAW 2024), Nov 2024, Amsterdam, Netherlands. pp.364-381, <10.1007/978-3-031-77792-9_22>.

Population : PeGazUs



Initialisation du graphe de faits (après étape 1)

0. Modélisation des données selon l'ontologie
1. Détection des **entités similaires**
2. **Tri temporel** des mentions
3. **Déduction des changements** liés aux entités (apparition, disparition)

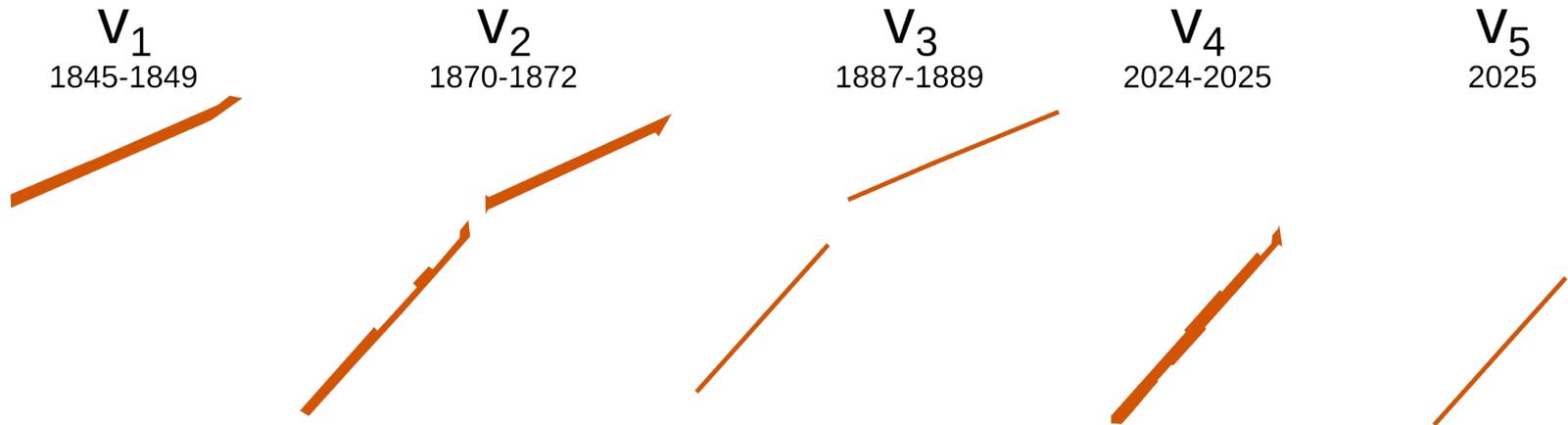
Population : Exemple

4. Traitement des attributs

C₁ "En 1857, elle est prolongée de la rue du Moulin-des-Prés à la rue Jonas."

C₂ Le "30 août 1978, sa portion orientale, [...], prend le nom de rue du Père-Guérin."

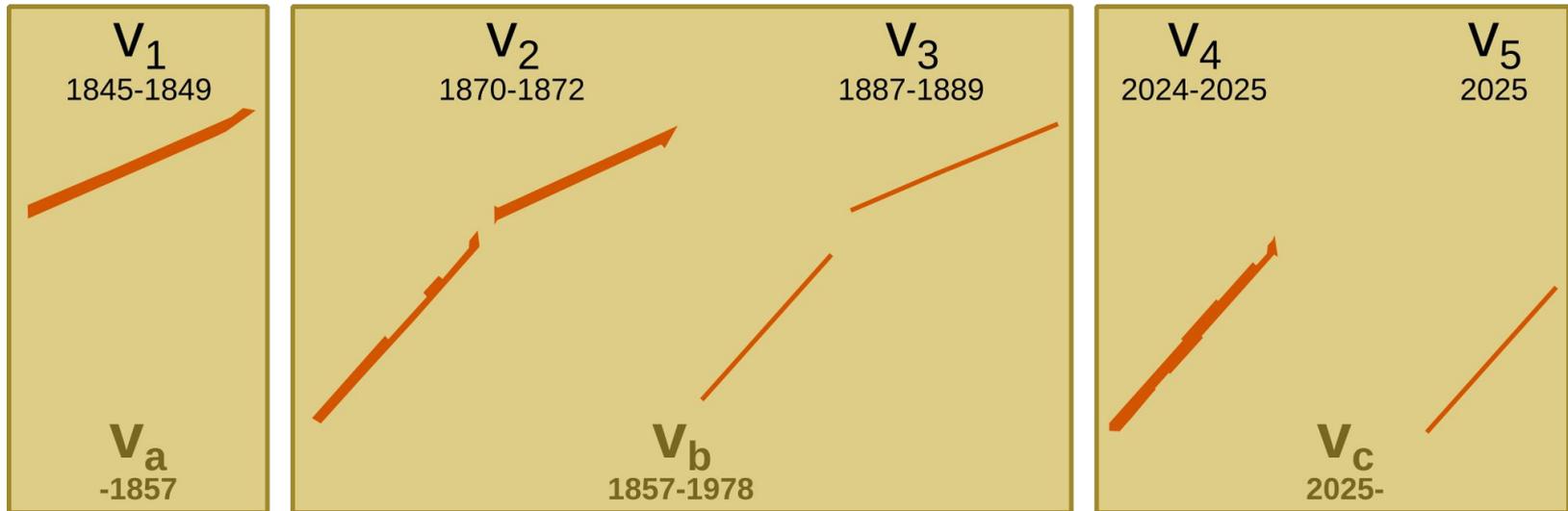
C₃ "La partie comprise entre les rues Bobillot et du Moulin des Prés a été dénommée rue du Père Guérin par arrêté municipal du 30 août 1978."



Population : Exemple

4. Traitement des attributs

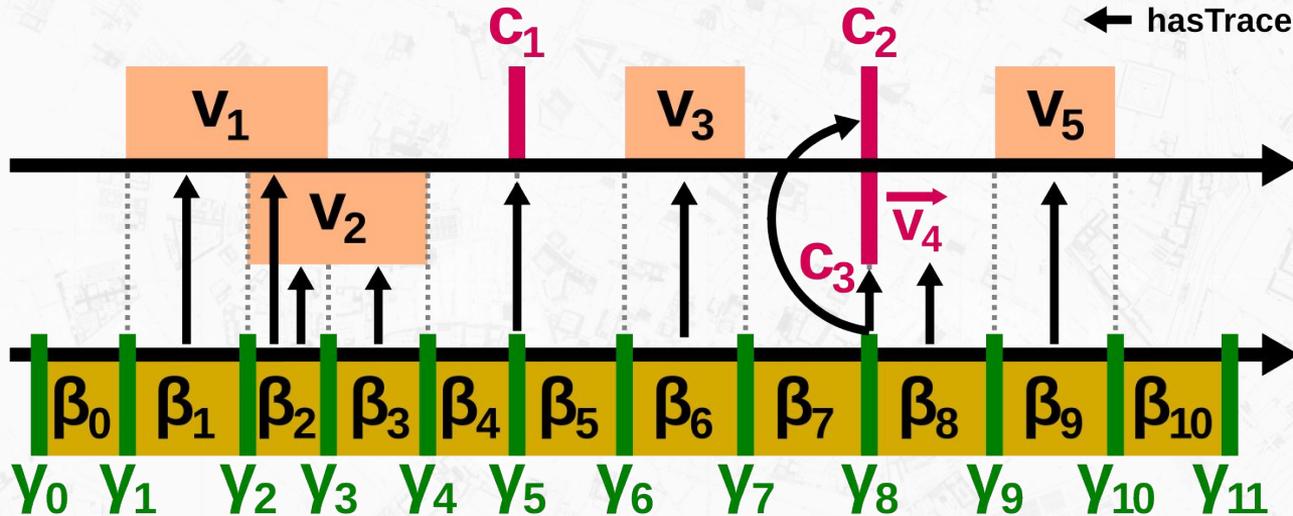
- C_1 "En 1857, elle est prolongée de la rue du Moulin-des-Prés à la rue Jonas."
 C_2 Le "30 août 1978, sa portion orientale, [...], prend le nom de rue du Père-Guérin."
 C_3 "La partie comprise entre les rues Bobillot et du Moulin des Prés a été dénommée rue du Père Guérin par arrêté municipal du 30 août 1978."



Ensemble des versions et changements liées à la géométrie de la rue Gérard

Population : PeGazUs

- Attributs : **fusion de versions** successives similaires :
 - Initialisation de **versions élémentaires** avec tri de ressources
 - Fusion de ces versions** selon l'agencement et leur valeur
- Création de changements et d'événements liés aux attributs

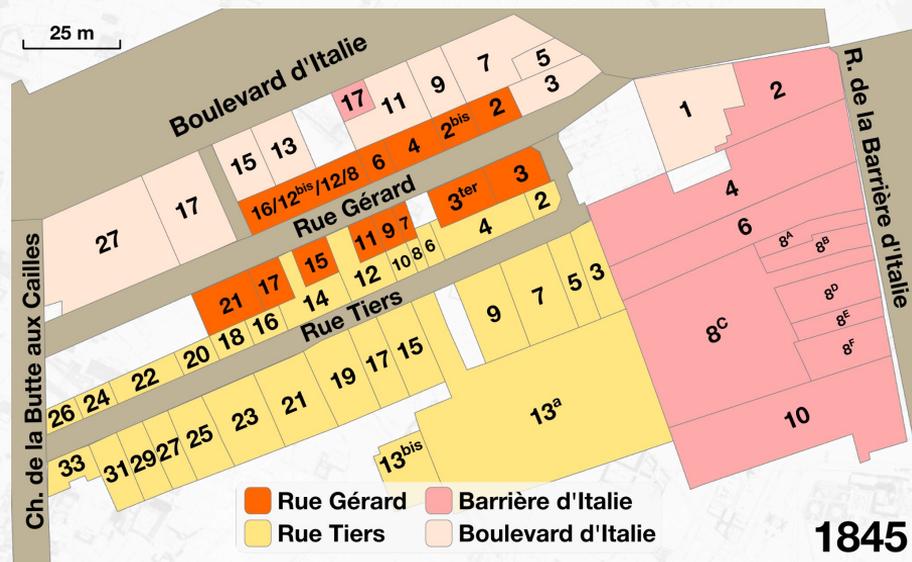


Exemple d'initialisation de versions élémentaires selon les données des sources

Évaluation

- **Vérité terrain inexistante**
- **Évaluation qualitative**
 - extraction de **snapshots** à différents instants
 - **étude individuelle** de l'évolution des entités
 - **cohérent géographiquement ?**

*Quelles voies existent en 1845 ?
Quelles étaient les adresses des parcelles en 1845 ?*



Snapshot du quartier de la Butte aux Cailles en 1845

Évaluation

Outil de visualisation de snapshots



Interface de visualisation affichant le territoire autour de la rue Gérard en 1888

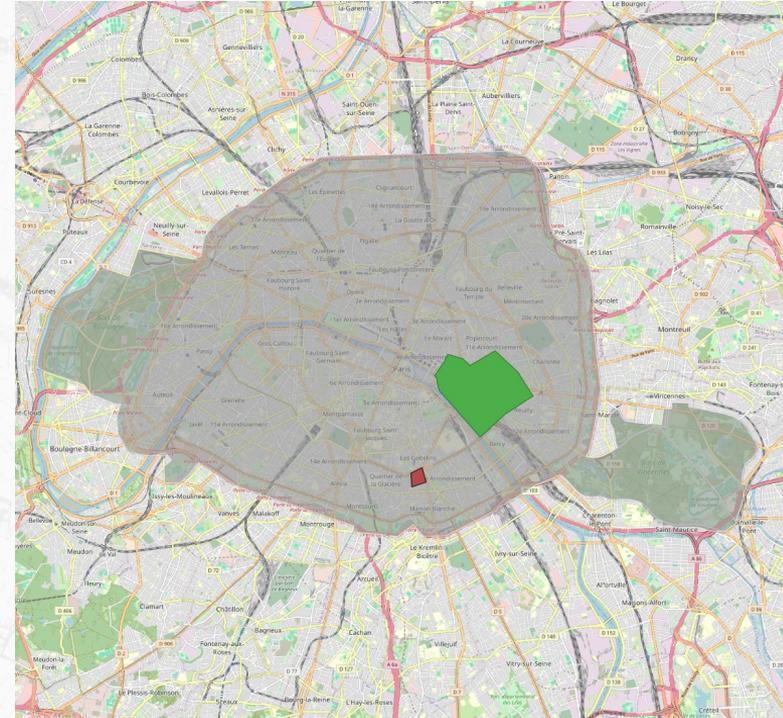
Contributions

Construction d'un répertoire géo-historique sous la forme d'un **graphe de connaissances**

- gestion de plusieurs **sources hétérogènes**
- traitement de données **fragmentaires et hétérogènes**
- intégration de données décrivant des événements

Perspectives

-  **Cas d'application**
 - travail sur un quartier plus grand
 - adaptation du code pour le temps de traitement
-  **Évaluation quantitative**
-  **Extraction et structuration des informations d'évolution des rues dans des corpus textuels à l'aide de LLM**



Localisation du nouveau quartier d'étude (en vert) comparée à la Butte aux Cailles (en rouge)



Merci de votre attention !

Ontologie, données et scripts disponibles

<https://github.com/umrlastig/pegazus-ontology>

<https://github.com/charlybernard/pegazus-extension>

charly.bernard@ensg.eu