



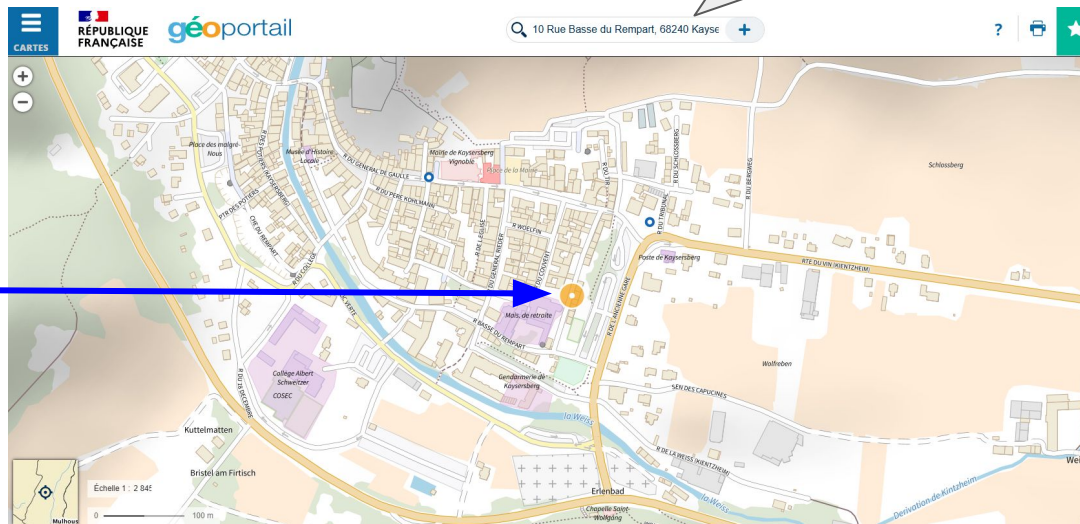
# Création de graphes de connaissances géohistoriques à partir de sources de données anciennes, hétérogènes et fragmentaires

[Nathalie ABADIE](#)

Séminaire SESAME - Montpellier  
12 décembre 2025

# Localiser des ressources, à la surface de la Terre et dans le temps

Hôtel Radix de Sainte-Foix,  
situé **10 Rue Basse-Du-Rempart**



Geoportail IGN, <https://www.geoportail.gouv.fr/>  
**Temps valide : 2025**

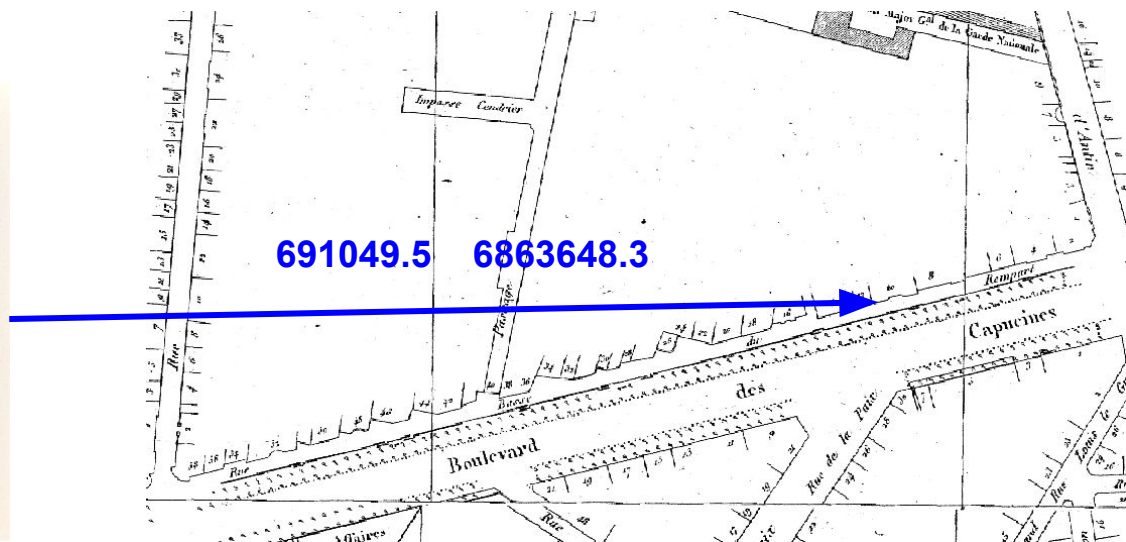
Par Jean-Baptiste Lallemand — Bibliothèque nationale de France, Domaine public,  
<https://commons.wikimedia.org/w/index.php?curid=13275379>

# Localiser des ressources, à la surface de la Terre

Hôtel Radix de Sainte-Foix,  
situé **10 Rue Basse-Du-Rempart**



Source gallica.bnf.fr / Bibliothèque nationale de France

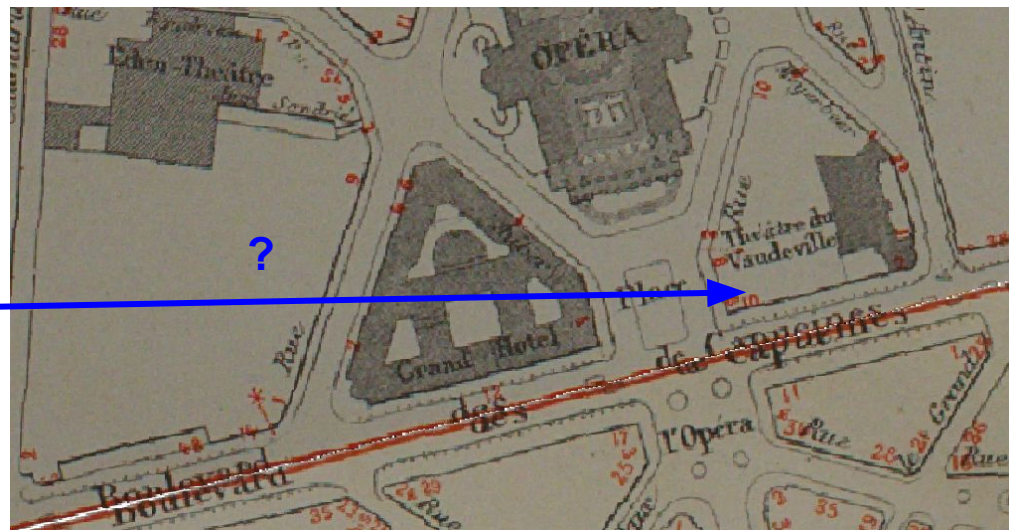
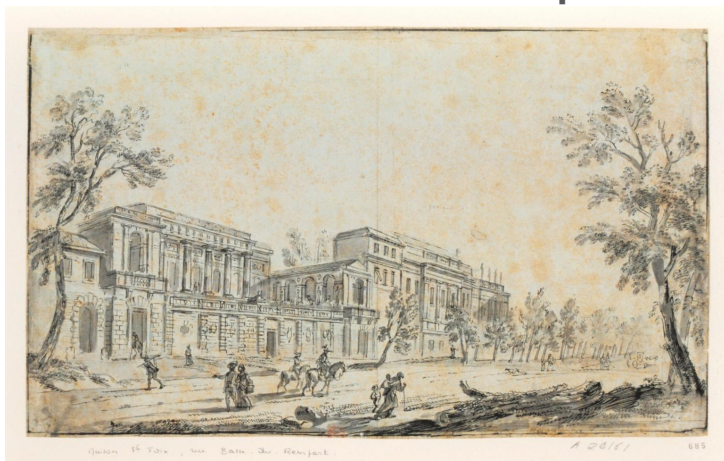


Par Jean-Baptiste Lallemand — Bibliothèque nationale de France, Domaine public,  
<https://commons.wikimedia.org/w/index.php?curid=13275379>

Atlas de Jacoubet géoréférencé,  
<http://geohistoricaldata.org/geoserver/wms>  
Temps valide: 1827 - 1836

## Localiser des ressources, dans le temps

Hôtel Radix de Sainte-Foix,  
situé **10 Rue Basse-Du-Rempart**



Par Jean-Baptiste Lallemand — Bibliothèque nationale de France, Domaine public,  
<https://commons.wikimedia.org/w/index.php?curid=13275379>

Atlas Municipal géoréférencé,  
<http://geohistoricaldata.org/geoserver/wms>  
**Temps valide : 1888**

# Quels référentiels pour localiser les ressources du Web?

Ensemble de données représentant l'état du territoire:

- homogènes,
- exhaustives,
- avec des références spatiales directes exactes, ou du moins, dont la qualité est connue, maîtrisée et documentée,
- avec des niveaux de détail variés (analyses à différentes échelles),
- à jour,
- disponibles à différentes dates, de façon aussi continue que possible,
- accessibles et interopérables.



# Limites des référentiels de données géographiques structurées

- Profondeur temporelle des référentiels de données géographiques structurées très faible (ex. Pour la BD TOPO IGN ~ début des années 2000).
- La logique de *snapshots* prévaut: pas ou peu de données spatio-temporelles.



2013



2015



2017

# Limites des référentiels de données géographiques images

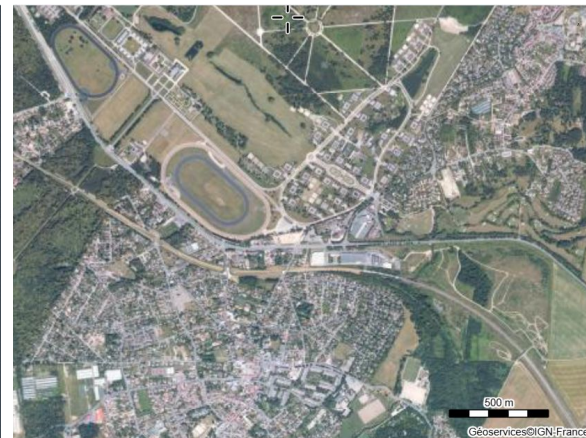
Environs de l'hippodrome de Boissy-Saint-Léger (Val-de-Marne)



Vers 1960



Vers 1980

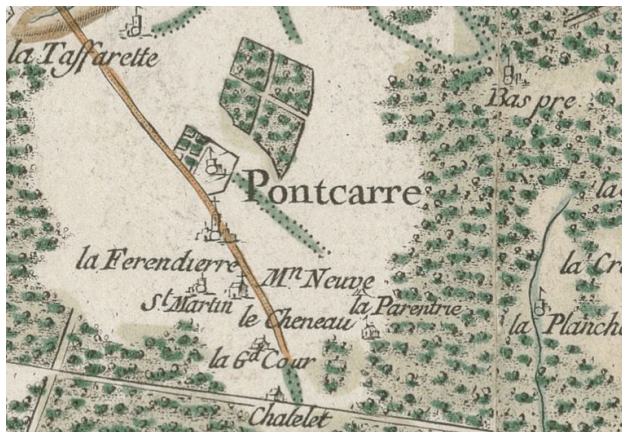


Vers 2000

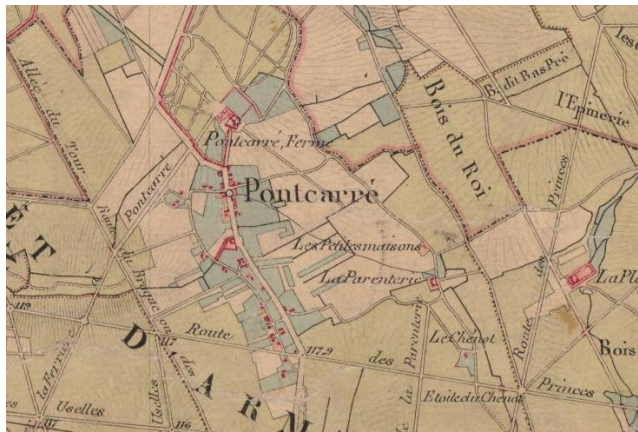
Images : Remonter le temps, IGN  
<https://remonterletemps.ign.fr/>

# Sources de données géographiques historiques: cartes anciennes

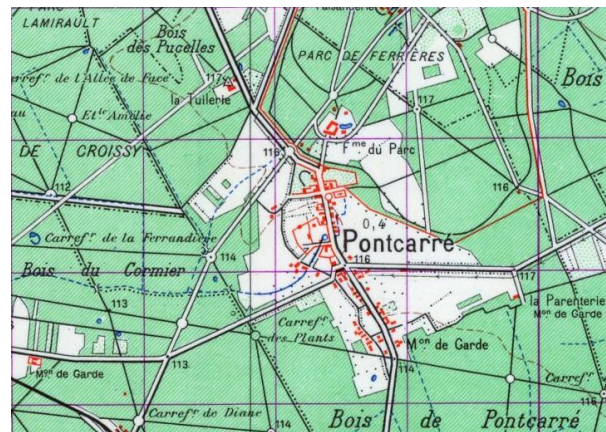
Environs de la commune de Pontcarré (Seine-et-Marne)



Carte de Cassini



Carte d'Etat-Major



Scan historique IGN 1950

Images : Remonter le temps, IGN  
<https://remonterletemps.ign.fr/>



CANTON, TRIAGES, ou Lieux dits.	NUMEROS DU PLAN provisoire déposé	O M S,		NATURES des Propriétés.	CONTENANCE	CONTENANCE	
		PROPRIETES Propriétés ou Usambles.	REMISES es		par	par nombre de culture.	
	+ 56	Bois de 1 <sup>re</sup> et 2 <sup>e</sup> qualité Puis 2 <sup>e</sup> qualité		Maisons et Cour.	32 15		60 M 2 <sup>e</sup> de
	+ 58	Bois		Forêt	37 60	37 60	2 <sup>e</sup>
	+ 56	Bois		Vignes, vergers	29 00	29 00	1 <sup>e</sup>

[illegible]

Source: Archives départementales du Val de Marne

<b>BAUDOUY</b> (place).	26 Longpré aîné, bijoutier en or et argent.	Bourgeois, fabr. de l'École communale de jeunes filles.
<b>EX AR., Hôtel-de-Ville.</b>	Saint-Amand, émailleur.	Berthelette, sœurs, vins.
Rue Tixerontière, pourtour	Gellier (A.), fab. cuivre-calc.	6 Verstaen, serrurier-mécanicien.
St Gervais, Saint-Antoine et	Bouillon (J.), bijoutier en or.	8 Michel, brossier.
Léonard Lebère.	Benoit, orfèvre-fabr. d'or.	9 Labottiere, serrurier.
<b>Lissoty (Vie), vins.</b>	40° Courtois, aîné, prop.	10 Sacrez, vins.
2 Privé, distillateur.	30 Bouton, fab. de cuir noir.	11 Baudouin, zinc.
Lemoine-Guzel et Leroy, au coin.	31 Desmarais, fab. bottes d'emballage.	13 Lejard, cloutiers et crépeux.
Chantrier, court.-garn.	32 Duvet, vins.	14 Lefebvre (Yve), fab. de

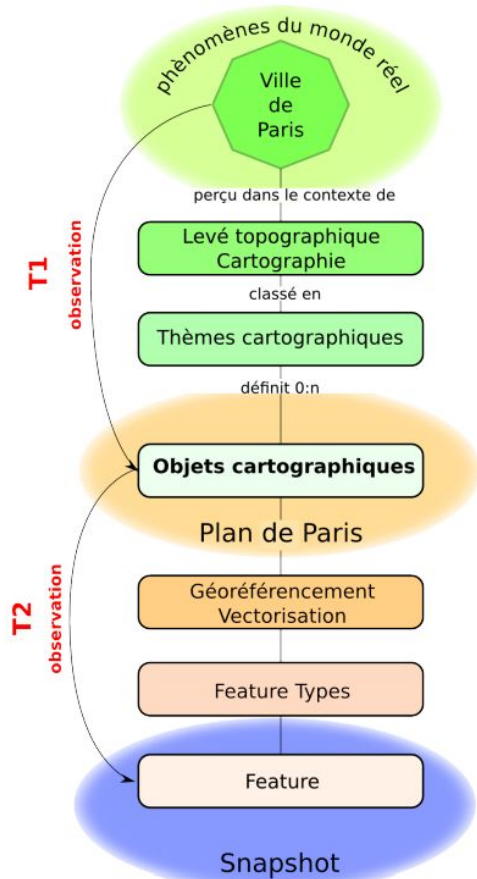
CENSIVE DE L'ÉVÊCHÉ.

*Adolphe Berty, Topographie historique du  
vieux Paris (1866-1897)*

[+ ajouter une déclaration](#)

Page Wikidata de la rue Louis-Philippe

# Limites des sources de données géographiques anciennes

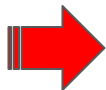


- Sources indirectes sur les états passés du territoire (à l'exception des photographies),
- Sources non structurées (textes ou images),
- Sources non géoréférencées,
- Sources hétérogènes: états et événements,
- Sources le plus souvent fragmentaires (spatialement ou temporellement).

# Des graphes de connaissances géohistoriques pour localiser les ressources du Web

Ensemble de données représentant l'état du territoire:

- homogènes,
- exhaustives,
- avec des références spatiales directes ~~exactes, ou du moins~~, dont la qualité est connue, maîtrisée et documentée,
- avec des niveaux de détail variés (analyses à différentes échelles),
- ~~à jour~~ **de temps valide connu**,
- disponibles à différentes dates, de façon aussi continue que possible, **avec des liens spatio-temporels entre les représentations des entités géographiques à différentes dates**,
- accessibles et interopérables.

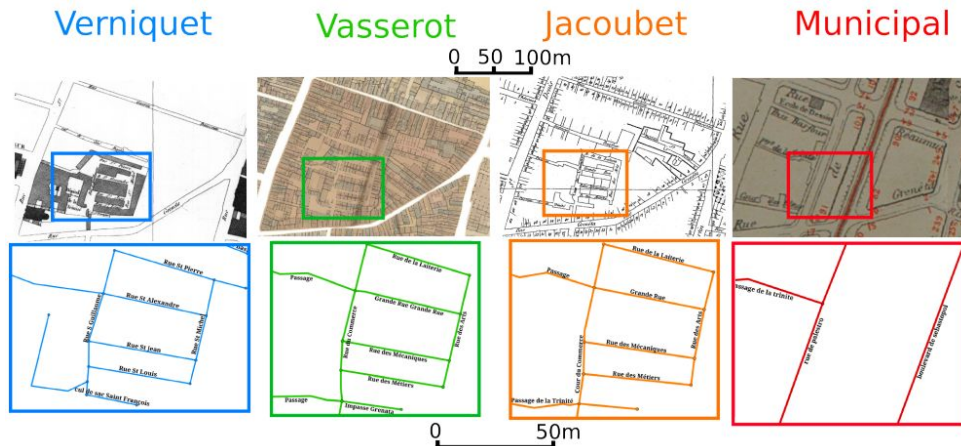
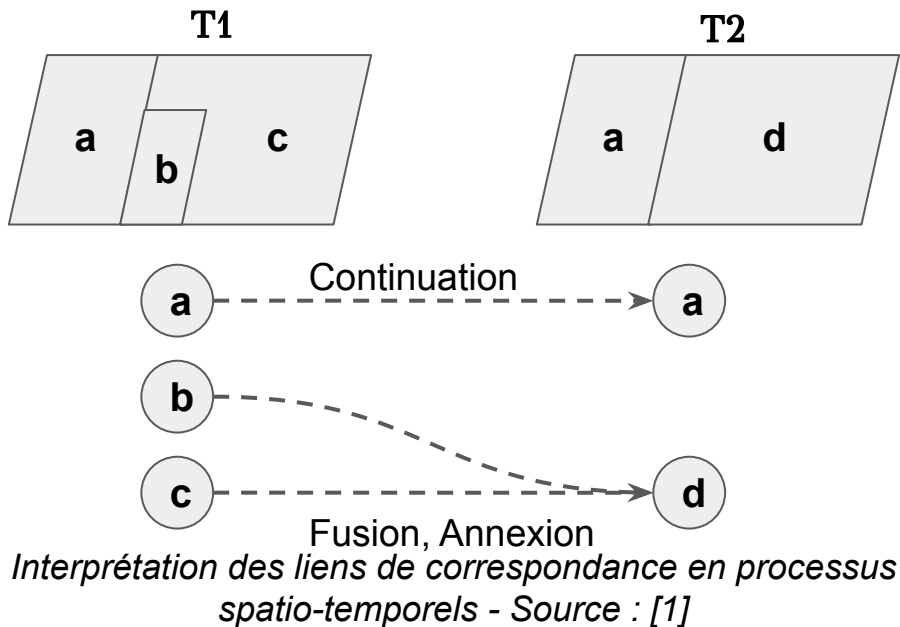


**Mettre à profit l'infrastructure du Web de données pour sortir des logiques de production de référentiels géohistoriques en silos.**

# Approches pour la création de graphes spatio-temporels

A partir de *snapshots* [1, 2, 3, 4] :

- Détection de relations de filiation entre entités géographiques par liage.
- Interprétation des graphes de filiation en processus spatio-temporels.



*États successifs de l'îlot de la Trinité et réseaux de rues vectorisés correspondants - Source : [2]*

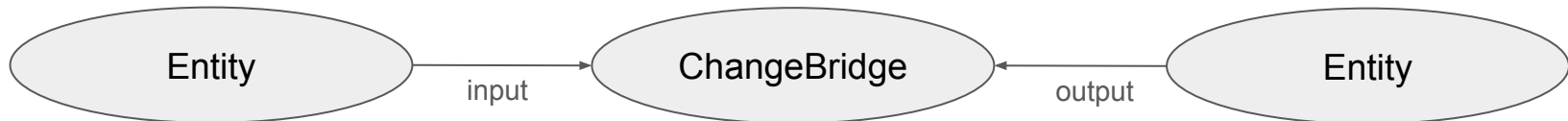


# Approches pour la création de graphes de connaissances géohistoriques d'unités territoriales

- Kauppinen et al. [5]: Municipalités en Finlande (depuis 1865)
- Territorial Statistical Nomenclature (TSN) [4]: Unités statistiques et administratives
- Hierarchical Historical Territory (HHT) [6]: Unité territoriales françaises avant 1789

Ces trois approches ont en commun de proposer:

- Une ontologie fondée sur le principe de Change Bridge.
- Un algorithme de peuplement qui infère les connaissances de changement (sauf [5] qui a déjà des données de changement préparées a priori)



# Limites pour représenter des entités géographiques qui évoluent dans le temps

**Les valeurs des propriétés des entités géographiques sont constantes sur leur période d'existence.**

→ Une commune ne peut pas changer de nom, une parcelle ne peut pas changer de type d'occupation du sol, etc.

**Les données utilisées sont déjà structurées, homogènes, normalisées, et désambiguïsées.**

→ Ne permet pas d'intégrer des données sur les événements qui modifient les entités géographiques et d'en déduire des connaissances sur leurs états;

→ Ne gèrent pas les questions de désambiguïsation des entités géographiques;

→ Ne gèrent pas l'intégration de nouvelles sources de données.

# Propositions

- Une ontologie qui applique le principe de Change bridge aux propriétés.
- Un algorithme de peuplement qui permet :
  - d'intégrer des données sur les états et les événements
  - d'intégrer de nouvelles sources
  - de désambiguïser les entités géographiques

→ Ces propositions proviennent des thèses de :

- **Solenn Tual**. Reconstitution automatique de la généalogie des parcelles et des bâtiments à partir de sources historiques. Soutenance prévue le 16/12/25.
- **Charly Bernard**. De données hétérogènes et fragmentaires provenant de sources multiples à un référentiel géohistorique d'adresses. Soutenance prévue (sous réserve d'autorisation) le 26/02/26.

Les illustrations présentées dans la suite en sont tirées.

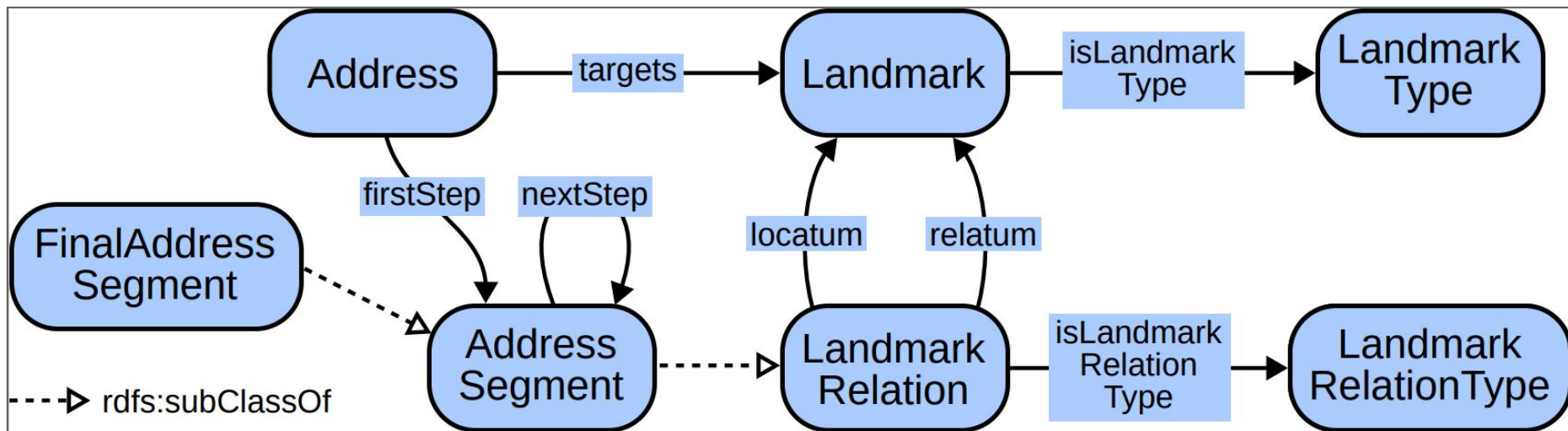
# L'ontologie PeGaZUs (PErpetual GAZetteer of approach-address UtteranceS)

- Conçue pour représenter des adresses et leur évolution au cours du temps:
  - Adresse conçue comme un cheminement d'un point de repère à l'autre dans la ville.
  - Reprend le principe du Change Bridge.
  - Ré-utilise des ontologies connues: OWL-Time et Location Core Vocabulary
- Développée avec SAMOD
- Ontologie, documentation et données de tests sont publiées sur un dépôt [11]:  
<https://github.com/charlybernard/pegazus-extension>

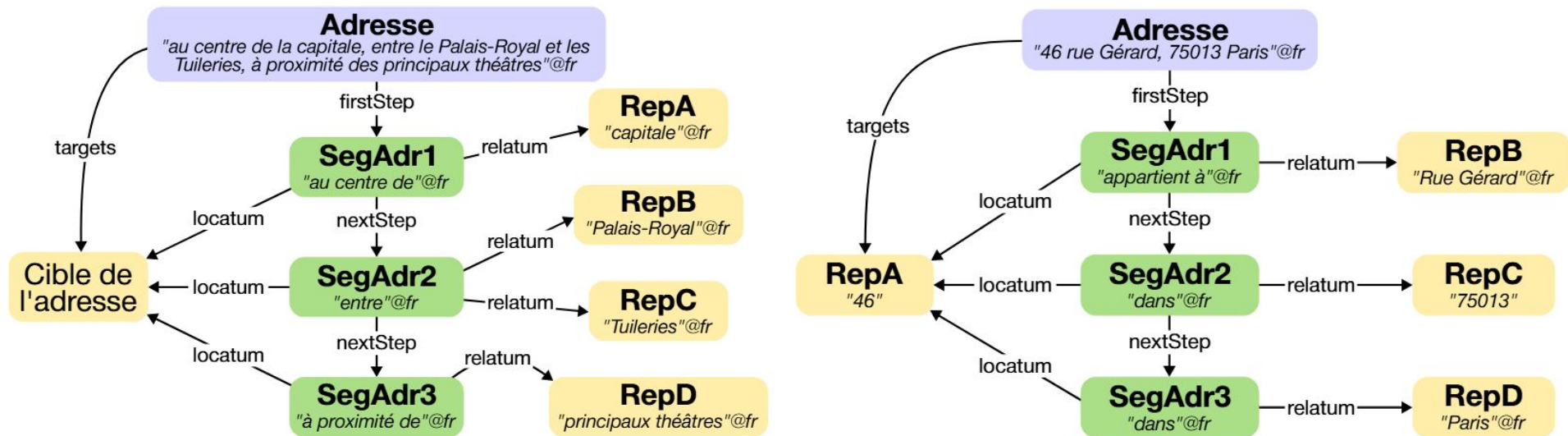


# PeGaZUs: des adresses modernes et anciennes

“ Un énoncé d'adresse est la référence spatiale indirecte d'un lieu. C'est 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.” [8]

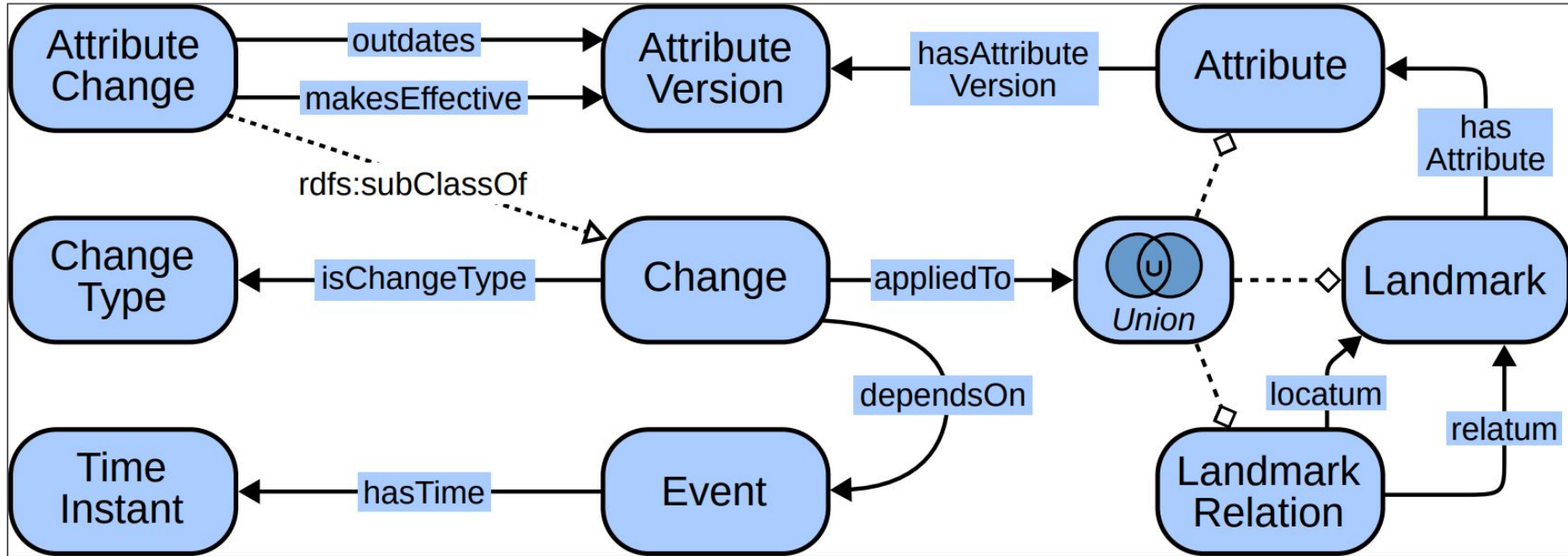


# PeGaZUs: des adresses modernes et anciennes

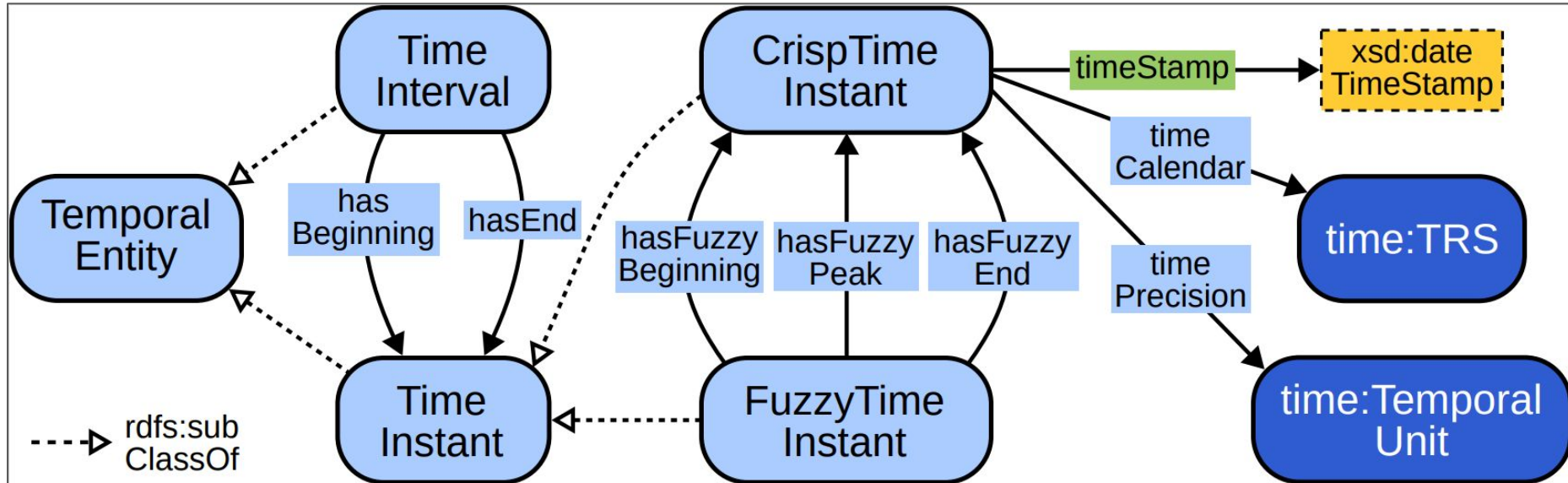


Exemples de représentations en graphe d'adresses ancienne et moderne selon l'ontologie PeGaZUs [8].

# PeGaZUs: des entités géographiques dont les propriétés évoluent

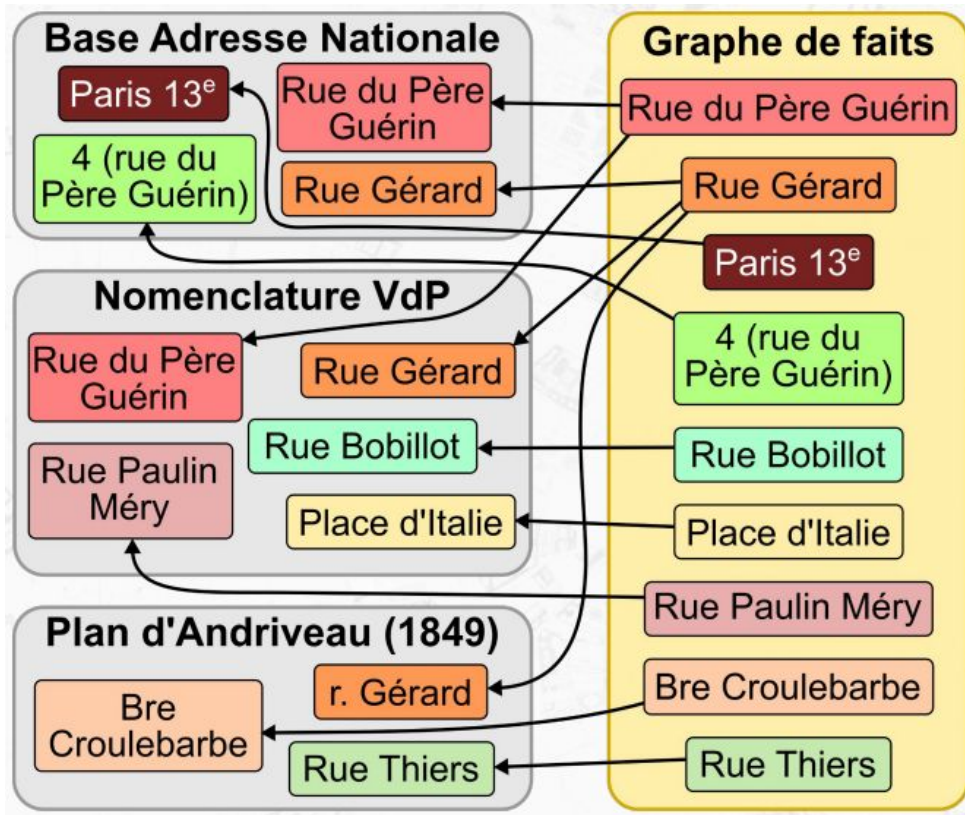


# PeGaZUs: des dates éventuellement incertaines





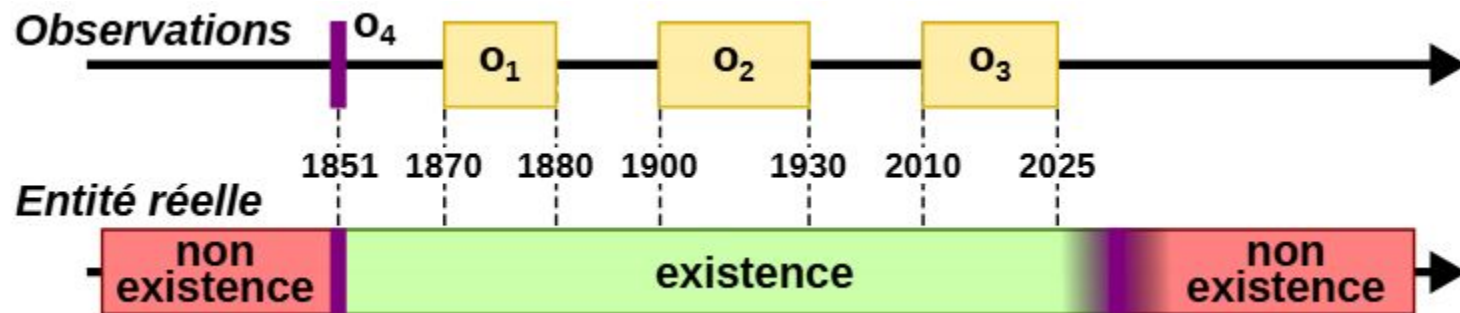
# Peuplement à partir de données fragmentaires



1. On crée un graphe de factoïdes conformes à l'ontologie pour chaque source d'observations et un graphe de faits contenant les entités provenant des graphes de factoïdes.
2. On lie les entités similaires entre graphes de factoïdes et graphe de faits à l'aide d'un critère de pseudo-identité et on propage leurs attributs, leurs relations, leurs changements et événements associés, etc.
3. On trie temporellement les factoïdes.

# Peuplement à partir de données fragmentaires

4. On crée les changements relatifs à l'apparition et la disparition des entités à partir des observations disponibles.

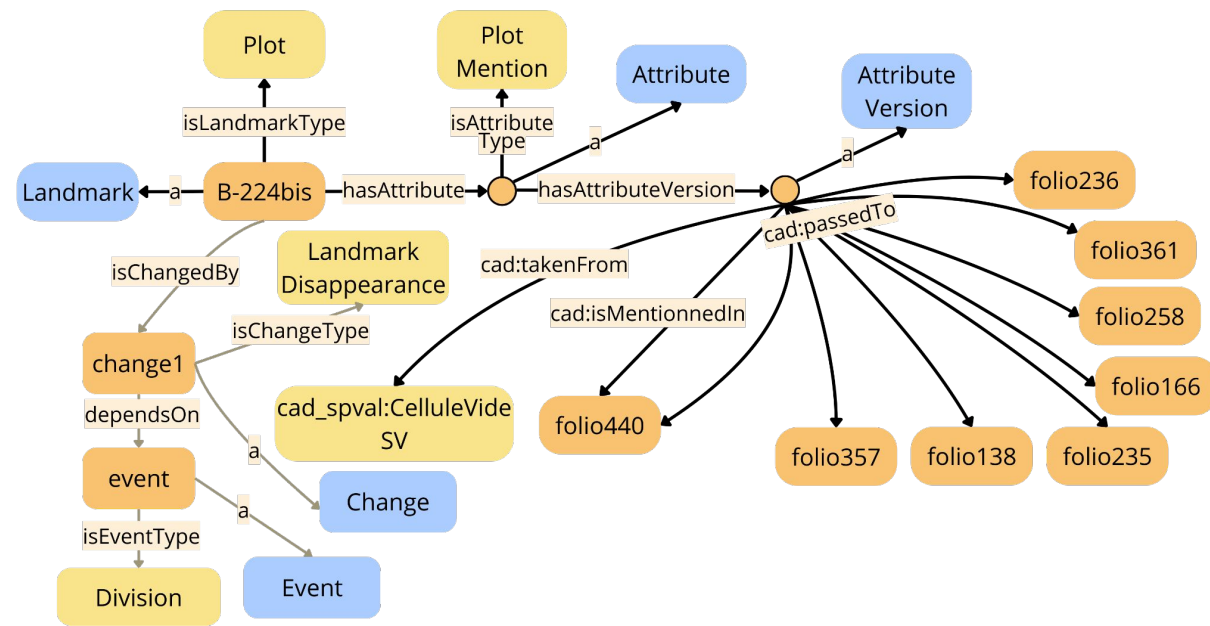


Reconstitution des états spatio-temporels de la rue Gerard à partir de ses observations. Source [8].

# Peuplement à partir de données fragmentaires

4. On crée les changements relatifs à l'apparition et la disparition des entités à partir des observations disponibles.

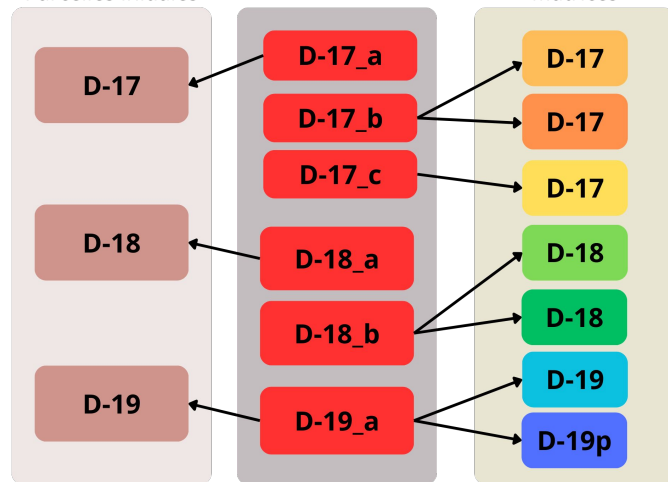
Cas du cadastre napoléonien [7]: On utilise l'ordre d'apparition dans les documents, les reports de folios, et les noms de propriétaires.



Graphe de  
factoïdes #1  
Parcelles initiales

**Graphe  
des faits**

Graphe de  
factoïdes #2  
Matrices

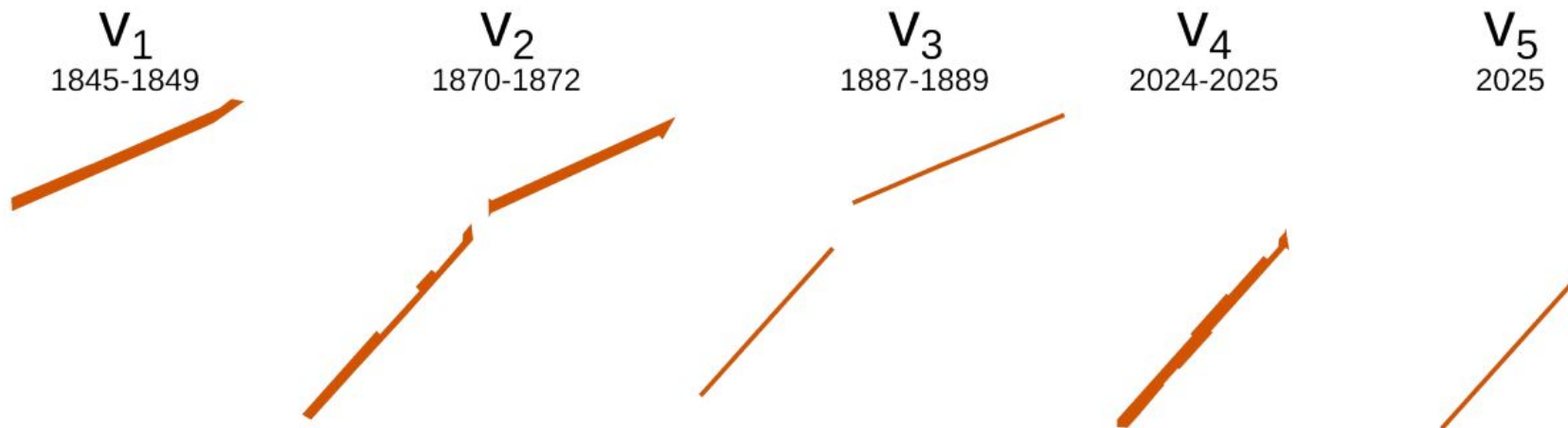


→ hasTrace

# Peuplement à partir de données fragmentaires

## 5.a. On reconstitue les évolutions élémentaires 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."

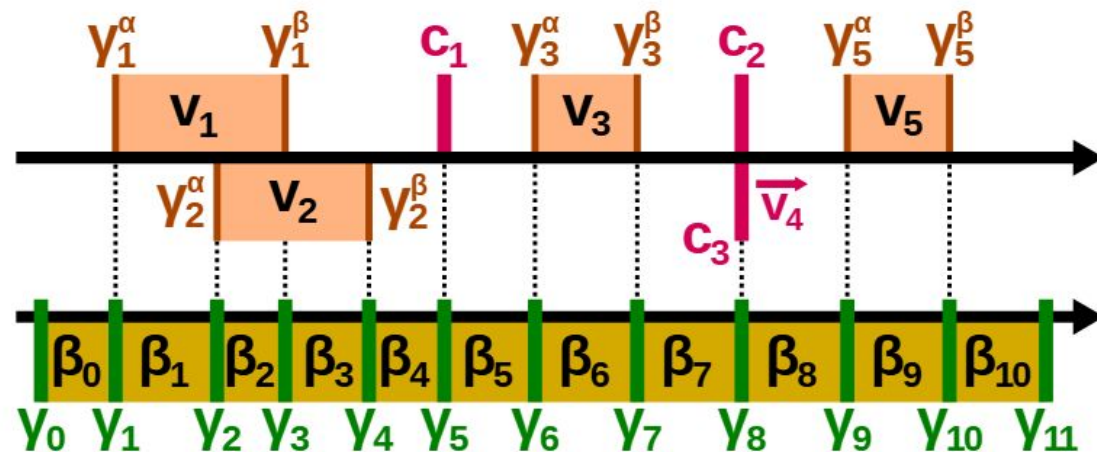


Observations disponibles pour la géométrie de la rue Gérard

Source: [8]

# Peuplement à partir de données fragmentaires

5.a. On génère et on ordonne temporellement les changements et les versions élémentaires des attributs à partir des observations de versions et de changements. On conserve un lien entre ces évolutions élémentaires et leurs observations d'origine.



$C_n$  changement

$V_n$  version avec temps valide

$\overrightarrow{V_i}$  version rendue effective par le changement

$Y_i$  changement élémentaire

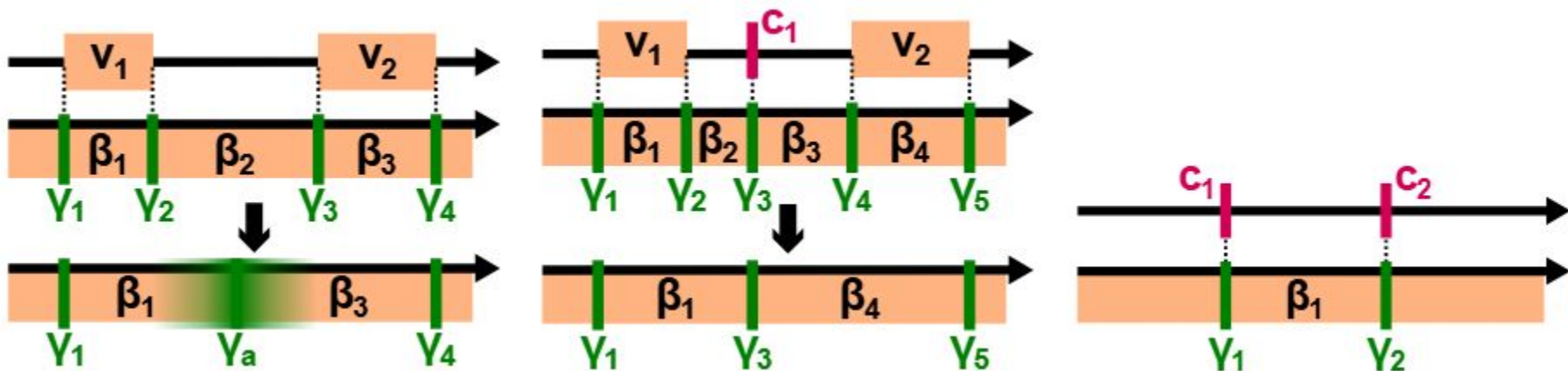
$\beta_j$  version élémentaire

$Y_k^\alpha$  changement fictif rendant effectif / périmant la version  $v_k$



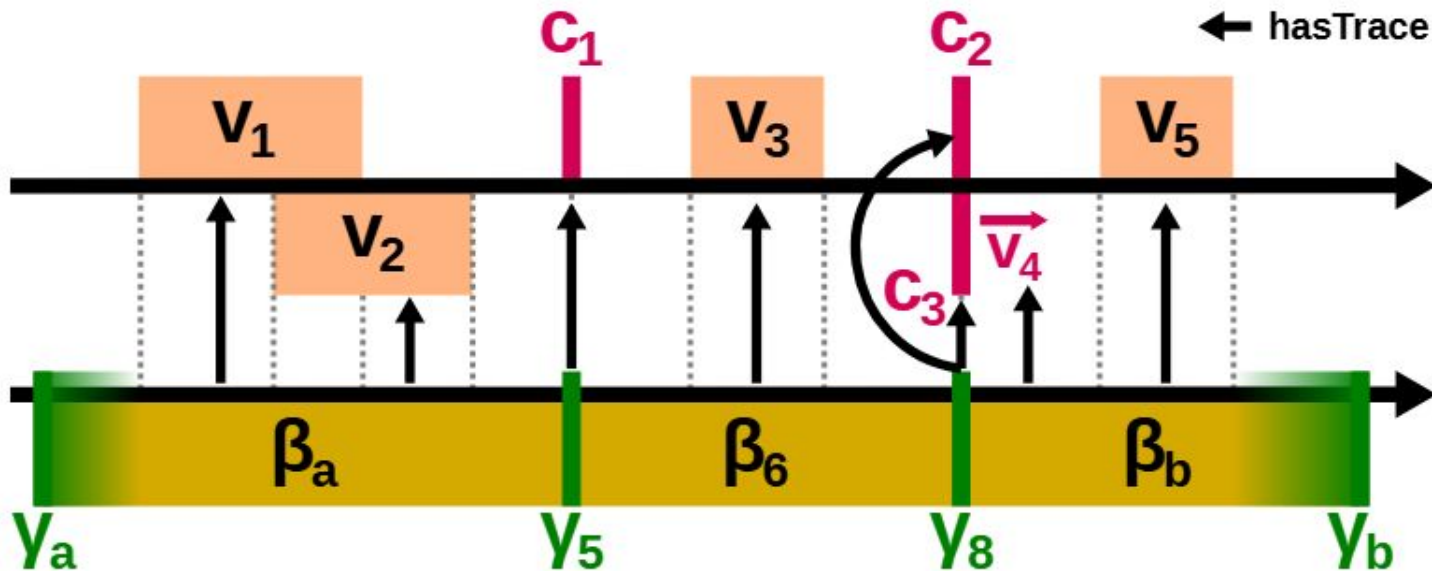
# Peuplement à partir de données fragmentaires

5.b. On agrège les évolutions élémentaires successives qui ne sont pas justifiées par une observation.



# Peuplement à partir de données fragmentaires

5.c. On agrège les versions élémentaires successives similaires et on crée les changements et événements associés pour produire une représentation dédoublonnée de l'évolution des attributs.



# Comment évaluer l'approche de peuplement?

→ Pas de vérité terrain et très difficile d'en faire une manuellement

→ Vérifier visuellement la cohérence géographique des résultats:

Test sur le quartier de la Butte aux Cailles, avec les données du cadastre napoléonien de Gentilly [7] [9].

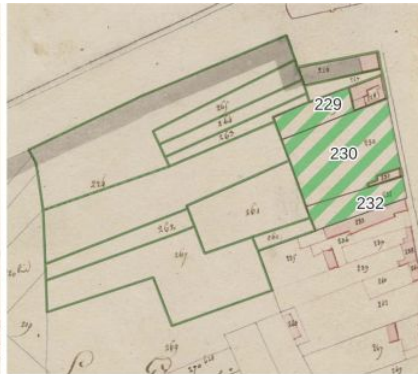
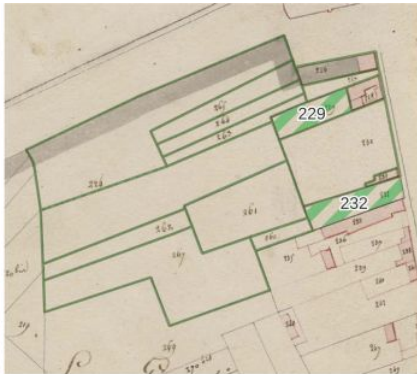
*Quelles sont les parcelles contenant un **jardin** (à différentes dates) ?*

1820

1830

1840

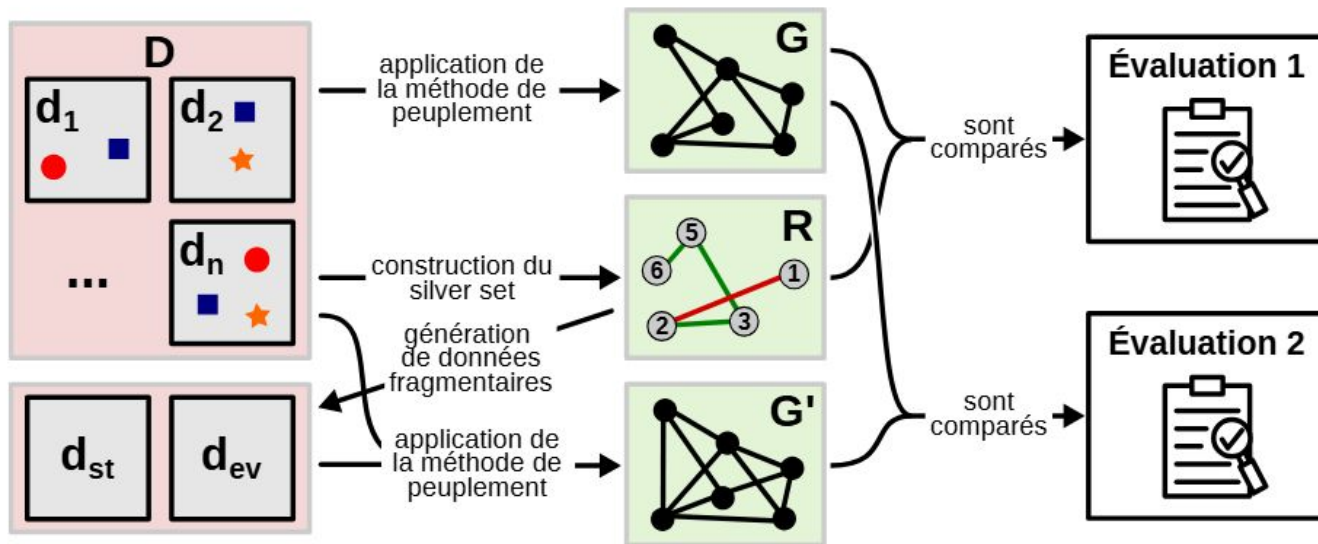
1850



# Comment évaluer l'approche de peuplement?

→ Tester l'approche sur des données pseudo synthétiques [8] [10] :

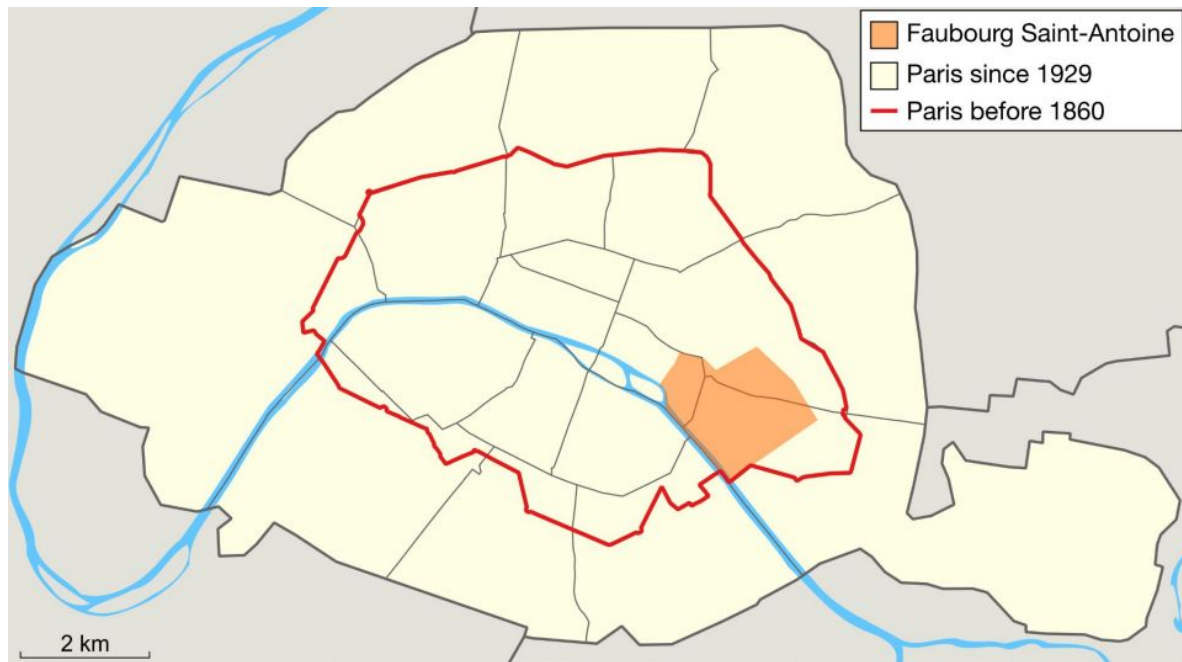
- Snapshots, utilisés pour générer un silver test set
- Snapshots, traités comme des fragments en entrée de l'algorithme
- Fragments, générés à partir des snapshot, et utilisés par l'algorithme



# Comment évaluer l'approche de peuplement?

## Snapshots (7874 adresses):

- cadastre de Paris (1806-1808) ;
- atlas Vasserot (1810–1836) ;
- atlas Jacoubet (1827–1839) ;
- atlas municipal de 1888 (1887-1889) ;
- la Base Adresse Nationale (2024) ;
- OpenStreetMap (2025).

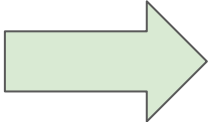




# Comment évaluer l'approche de peuplement ?

Evaluation 1 :

$$S(R, G) = \frac{1}{|A|} \sum_{a \in A} s(a)$$

 0.998

$s(a) = 1$  ssi les adresses ont:

- les mêmes intitulés,
- les mêmes temps valides,
- les mêmes versions de géométries (valeurs + temps valides).

Evaluation 2 :

	$S_{\text{stru}}(G, G')$	$S_{\text{temp}}(G, G')$
$G' = G_{\text{st}}$	1	1
$G' = G_{\text{st+ev}}$	1	0,1330

# Conclusion & perspectives

On peut générer des graphes de connaissances géohistoriques à partir de données anciennes, hétérogènes, ambiguës et fragmentaires.

- Intégrer plus de données d'événements (extraites de textes)
- Reconstruire les géométries des parcelles cadastrales
- Reconstruire la généalogie du bâti
- Détecter & corriger les incohérences entre sources
- etc.

# Références

- 1) Del Mondo Géraldine. Un modèle de graphe spatio-temporel pour représenter l'évolution d'entités géographiques. Thèse de doctorat, Université de Brest, 2011.
- 2) Duménieu Bertrand. Un système d'information géographique pour le suivi d'objets historiques urbains à travers l'espace et le temps. Thèse de doctorat, Ecole des Hautes Etudes en Sciences Sociales, 2015.
- 3) Aurélie Leborgne, Adrien Meyer, Henri Giraud, Florence Le Ber, and Stella Marc-Zwecker. Un graphe spatio-temporel pour modéliser l'évolution de parcelles agricoles. In Conférence internationale francophone en analyse spatiale et géomatique SAGEO, 2019.
- 4) Camille Bernard, Marlène Villanova-Oliver, and Jérôme Gensel. Theseus : A framework for managing knowledge graphs about geographical divisions and their evolution. Transactions in GIS, 2022.
- 5) Kauppinen, T., Väättäinen, J. and Hyvönen, E., 2008, June. Creating and using geospatial ontology time series in a semantic cultural heritage portal. In European Semantic Web Conference (pp. 110-123). Berlin, Heidelberg: Springer Berlin Heidelberg.
- 6) Charles, W., Aussenac-Gilles, N. and Hernandez, N., 2023, May. HHT: an approach for representing temporally-evolving historical territories. In European Semantic Web Conference (pp. 419-435). Cham: Springer Nature Switzerland.

## Références... suite

- 7) **Solenn Tual**. Reconstitution automatique de la généalogie des parcelles et des bâtiments à partir de sources historiques. Thèse de doctorat de l'Université Gustave Eiffel. Soutenance prévue le 16/12/25.
- 8) **Charly Bernard**. De données hétérogènes et fragmentaires provenant de sources multiples à un référentiel géohistorique d'adresses. Thèse de doctorat de l'Université Gustave Eiffel. Soutenance prévue (sous réserve d'autorisation) le 26/02/26.
- 9) Charly Bernard, Solenn Tual, Nathalie Abadie, Bertrand Duménieu, Julien Perret, Joseph Chazalon. 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.  
[https://hal.science/hal-04721538/file/EKAW2024\\_camera\\_ready%20%281%29.pdf](https://hal.science/hal-04721538/file/EKAW2024_camera_ready%20%281%29.pdf)
- 10) Charly Bernard, Nathalie Abadie, Bertrand Duménieu, Julien Perret. Reconstructing the Temporal Evolution of Geographic Entities from Fragmentary Knowledge. International Knowledge Graph and Semantic Web Conference, Nov 2025, Leipzig, Germany. <https://hal.science/hal-05308831>
- 11) Charly Bernard, Nathalie Abadie, Julien Perret, Bertrand Duménieu. Création d'un référentiel géo-historique d'adresses à partir de sources multiples. Atelier GAST - Gestion et l'Analyse de données Spatiales et Temporelles de la conférence Extraction et Gestion de Connaissances, Jan 2024, Dijon, France. <https://hal.science/hal-04490732>