

IGN

PeGazUs

A knowledge graph based approach to build urban perpetual gazetteers

Charly Bernard¹, Solenn Tual¹

Nathalie Abadie¹, Bertrand Duméniou², Julien Perret¹, Joseph Chazalon³

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

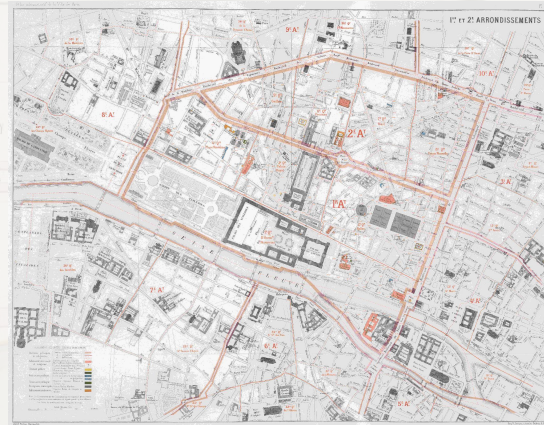
² CRH, EHESS, Paris, France

³ LRE, EPITA, Le Kremlin-Bicêtre, France



Context

- Growing volume of **digitized archival records** available online
 - ⇒ significant amount of **fragmentary** geohistorical information
- Sources need to be **indexed** both **spatially and temporally** to enhance information retrieval.
 - ⇒ very useful for research in humanities and for GLAM



Administrative atlas of the districts of Paris, 1868

Valat, vins en gros. pass. du Bureau, 23.
Valat fils, vins et hôtel, pl. de la Nation, 21, et
av. du Trône, 1.
Valax, corsets, r. Auber, 9.
Valax, tailleur, r. Boudreau, 2.
Valaz, vins, r. Rochambeau, 10.

AMADOU ET ALLUMETTES. — Pour les ALLUMETTES OXIGÉNÉES.
Voyez BRIQUETS PHYSIQUES.

DARRAS (Thomas), r. de la Vieille-Monnaie, 10.	GALLIENNE jr., r. de la Heaumerie, 3.
Briques et veilles, mèches à quinquets, mèches souffrées, pierres, agarie de chêne, liège en planches, bouchons.	Brûle-tout, boîtes à briquet, mèches à vin et veilles, mèches souffrées, pierres, agarie, bouchons, liège.
	LEROY, r. Aubry-le-Boucher, 43.

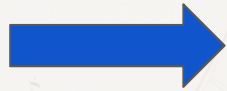
Extracts of 19th century Paris trade directories

Motivations

Indexation requires to have **perpetual gazetteers** that describe landmarks located at **any coordinates** at **any date** including:

- different names
- properties
- sources that assert their existence
- evolution

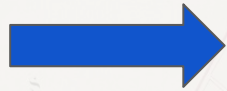
Challenge



Build this perpetual gazetteer

Representation and **integration of fragmentary data**
extracted from several sources (diverse and heterogeneous)

Challenge



Build this perpetual gazetteer

Representation and **integration of fragmentary data** extracted from several sources (diverse and heterogeneous)

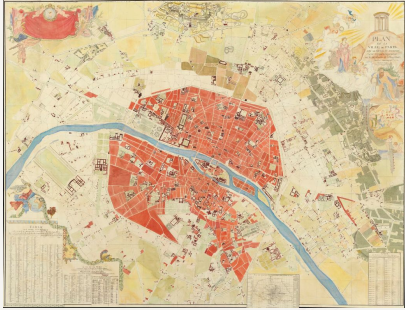
→ **Knowledge graphs** are a suitable solution!

Research questions

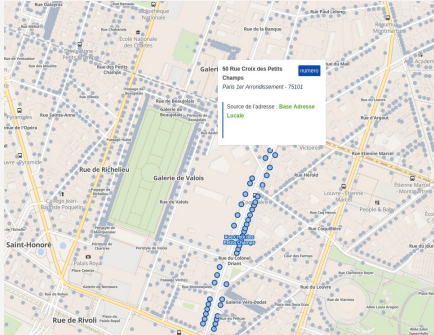
1. Which **ontologies** are suitable to build perpetual gazetteers?
 - structure named places that are useful to localize resources
 - retrieve all named places that refer to the location at any date
 - preserve the link between data and sources
2. How to **populate an ontology** of historical landmarks?
 - get the evolution of their attributes
 - get the evolution of their relations with other features

Data

Sources describe addresses and geographical entities over time



Verniquet, Plan de Paris (1791)



Base Adresse Nationale (2023)

rue Louis-Philippe (Q3448913)

Élément [Discussion](#)

voie de Paris modifier

[Plus de langues](#)
Configurer

Langue	Libellé	Description	Egalement connu comme
français	rue Louis-Philippe	voie de Paris	

Toutes les langues saisies

Déclarations

nature de l'élément 0 rue modifier
0 référence
[+ ajouter une référence](#)
[+ ajouter une valeur](#)

partie de 0 anciennes voies de Paris modifier
0 référence
[+ ajouter une référence](#)
[+ ajouter une valeur](#)

pays 0 France modifier
0 référence
[+ ajouter une référence](#)
[+ ajouter une valeur](#)

localisation administrative 0 Paris modifier
0 référence
[+ ajouter une référence](#)
[+ ajouter une valeur](#)
[+ ajouter une déclaration](#)

Wikidata page of Rue Louis-Philippe

Bégault, *courtier gourmet*, q. Bourbon, 49.
Begin, *Calandre*, 19.
Begis, *f. colle-sortie et huile de pieds de bœuf*, Ormeaux, 2,
pl. du Trône.
Béglé, *cordonnier*, Jacob, 56.
Béglet et co., *commiss. draps*, Thibautodé, 8.

BAUDOYER (place). IX Arr. <i>Hôtel-de-Ville</i> . Rue Texaenderie, pourtour St-Gervais, Saint-Antoine et Renard-Lefèvre.	25* Longpré aîné, <i>bijoutier en or et argent</i> . Saint-Omer, <i>émailleur</i> . Cellier (A.), <i>graveur-ci- scieur</i> . Bousseau (J.), <i>bijoutier en or</i> . Beuot, <i>orfèvre-fabr.</i> Lérisy, <i>denteur</i> . 30 Bouton, <i>fab. de cuir ver- nis</i> . Chantrier, <i>court.-gourm.</i>	Bourguille, <i>fabr. de presses</i> . Vautain, <i>passementier</i> . Finino jnc, <i>bronce doré</i> . Balé aîné, <i>fabr. de val- lons</i> . Gaulin, <i>chapetier</i> . Moisy, <i>tabletter</i> . 49* Cendrier aîné, <i>prop.</i> Desmarests, <i>fab. bottes d'emballage</i> . Ferrand, <i>lapidaire</i> .	7 Ecole communale de jeu- nes filles. Berthelot, <i>vins</i> . 6 Verstaen, <i>scrrurier-mé- canicien</i> . 8 Michel, <i>brossier</i> . 9 Labottiere, <i>serurier</i> . 10 Sacrez, <i>vins</i> . 12 Baudouin, <i>épici.</i> 15 Lejard, <i>clouteries et cré- pins</i> . 14 Baduel (Yve), <i>fab. de</i>
---	---	--	---

Commercial directories of Paris during XIXth century

346

TOPOGRAPHIE HISTORIQUE DU VIEUX PARIS.

situation, à la rue de ce nom et à celle de la Montagne-Sainte-Genève, qui seront décrites dans le volume suivant : RÉGION ORIENTALE DE L'UNIVERSITÉ.

CENSIVE DE L'ÉVÊCHÉ.

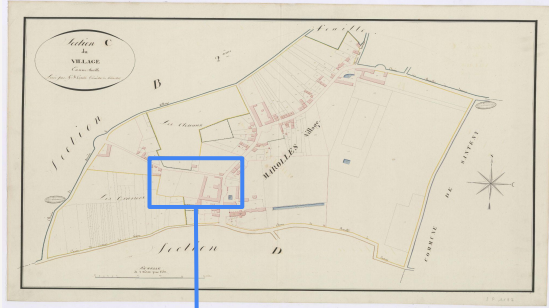
MAISON DE LA COURONNE (1575), faisant le coin occidental de la rue des Carmes. Elle est dite, en 1322, LA MAISON DE LA GRANT COURONE et DE LA PETITE COURONE, « au coin de la ruë des Noïers et de la ruë Saint Hilaire »; et en 1553, « maison devant les Carmes, faisant angle de la rue Saint-Hilaire », laquelle maison fut depuis divisée en trois corps de logis : LA COURONNE, L'IMAGE SAINT MARTIN et LA BOUTEILLE⁽¹⁾.

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

Data

Archival records: the Napoleonic Land Registry (1808-1950)

Cadastral index map (1810)



unique graphic representation of plots

legend of

handwritten spatial temporal referential of plots

Initial register (1812)

CANTON, TRIAGES, ou Lieux dits.	NUMÉROS DU PLAN		NOMS, PROFESSIONS, DEMEURÉS des Propriétaires ou Usufruitiers.	NATURES des Propriétés.	CONTENANCE	CONTENANCE	
	provisoire	défini			par Propriété.	par nature de culture.	
	+	54	Beethier, alex. M ^{re} Prince de Neufchâteau	Maison et Cour	30 15		00. M. 2 ^{de} date
	+	55	Idem	Jardin	37 60	37-00	2 ^e
	+	56	Idem	Vign. et bois	29 00	29 00	1 ^e

Description of the plots C-54, C-55 and C-56 at the creation of the land registry

Change registers (here: 1812-1822)

ARTICLES.	NOMS, PRÉNOMS, PROFESSIONS ET DEMEURÉS des Propriétaires et Usufruitiers.	INDICATION				NOMBRE D'ARPENS, PERCHES ET MÈTRES,		CLASSES.	REVENU		REVOIS pour les Mutations.	
		de la section.	du N° de la section.	de CANTON, triage ou lieu dit.	DE LA NATURE de la propriété.	par article de SECTION.	par article DE MATRIÈRE.		de chaque article de SECTION.	de chaque article DE MATRIÈRE.	tiré de l'article.	porté à l'article.
9	Beethier, alexandre Prince de Neufchâteau à Gros Bois	B	90	chef-lieu	Commune	arp. p. m.	arp. p. m.		fr. c.	fr. c.		
						2 06 99	2	99 99				76
						0 11 07	1	11 07				
						0 30 16	1	30 60				
						0 37 60	2	37 70				
						0 29 00	1	26 67				

Versions of the plots C-54, C-55 et C-56 at a given date (snapshot)

State of the art

Existing approaches

- Kauppinen et al. [1]: Finnish municipalities (since 1865)
- Territorial Statistical Nomenclature (TSN) [2]: administrative and statistical units
- Hierarchical Historical Territory (HHT) [3]: territorial subdivisions of France before 1789

⇒ **Change-Bridge** concept [1]: describes changes between landmark versions

⇒ Treat only landmarks versions **which attributes never evolve**

⇒ Used data: **structured, normalized, aggregated and interpreted**

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

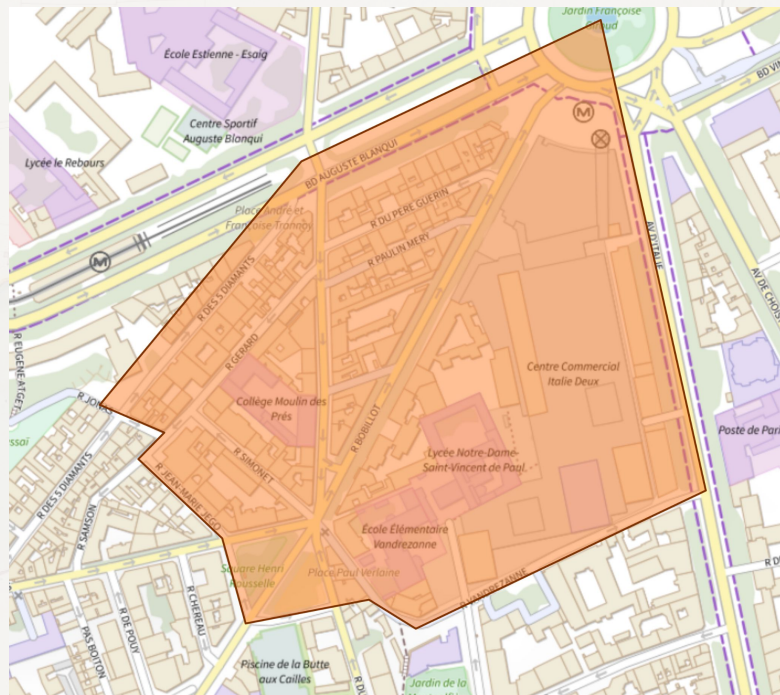
[2] Bernard et al. (2018) Modeling changes in territorial partitions over time: Ontologies tsn and tsn-change

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

Contributions

PeGaZus : Perpetual Gazetteer of approach-address UtteranceS

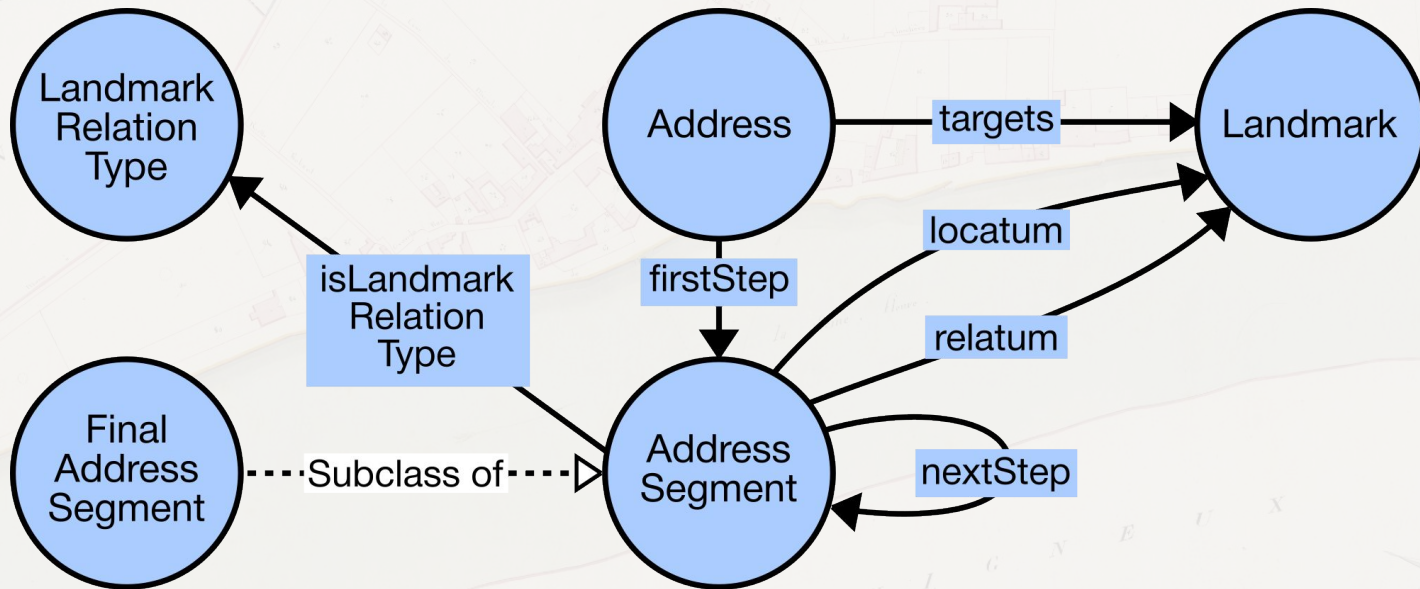
- **Ontology** represents:
 - addresses and landmarks,
 - with **attributes that evolve** over time
- **Population method** retrieves evolutions of addresses, plots and of their attributes
- Applied on a whole neighborhood in Paris (France) since the 1800s
- Map-based evaluation



Delimiting the Butte aux Cailles neighborhood in Paris, France

Ontology

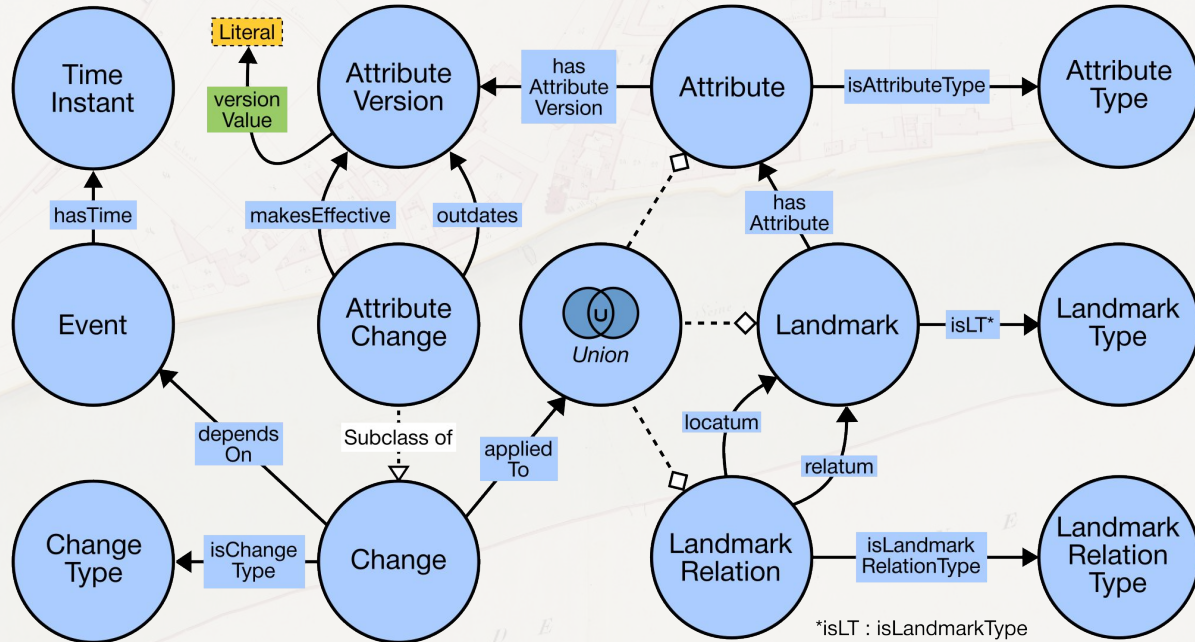
Address: statement which targets a place described as an ordered sequence of landmark relations [4]



Part of the ontology on addresses

Ontology

Evolution: representation of events. An event is related to an evolution on the territory, may imply multiple changes on data.

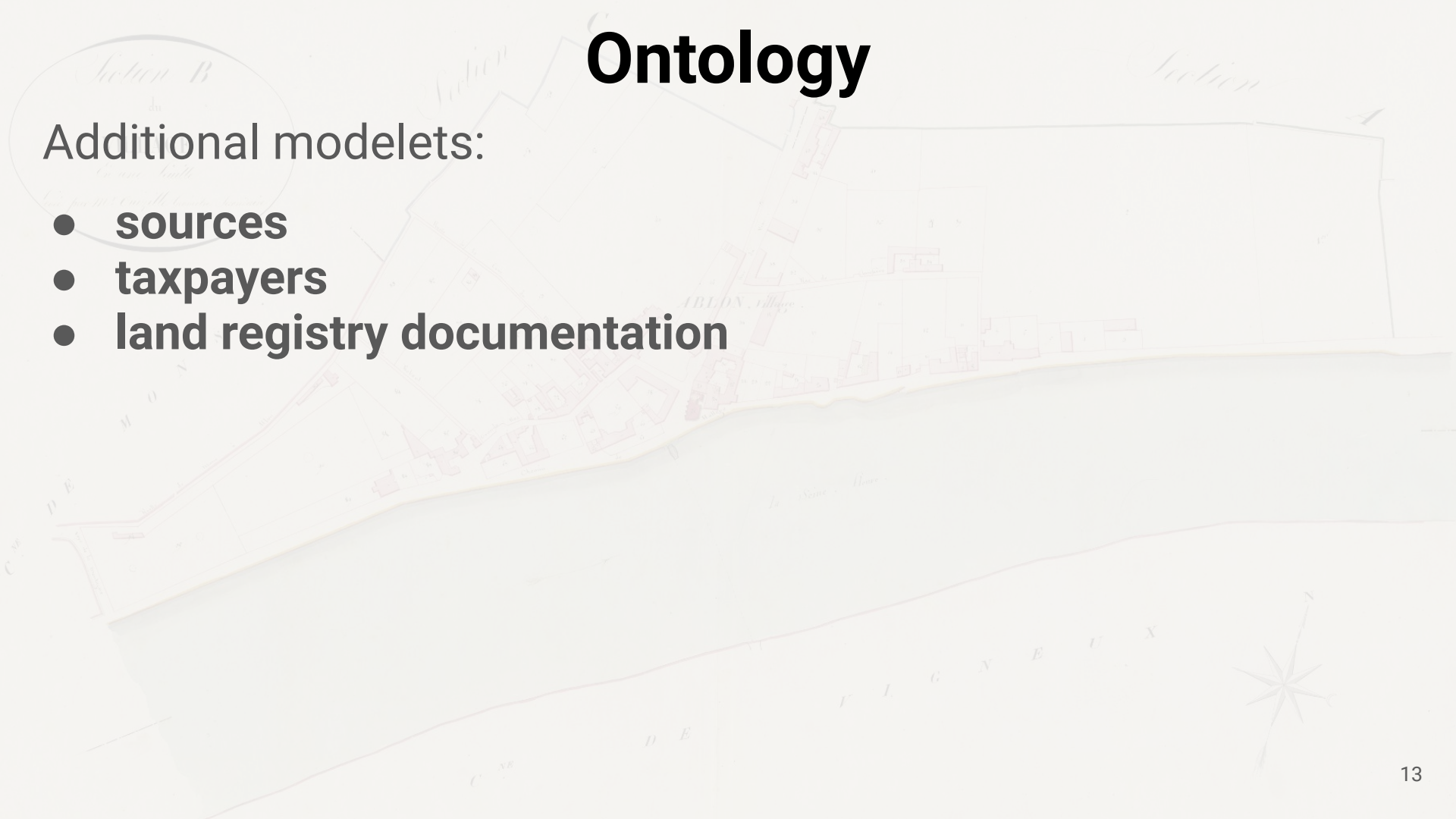


Part of the ontology on temporal evolution

Ontology

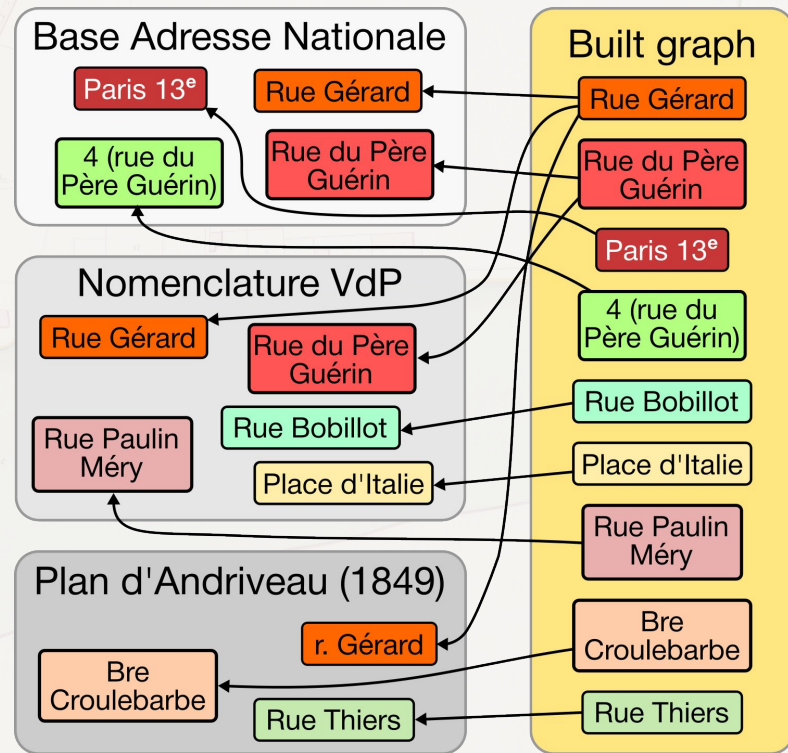
Additional modelets:

- **sources**
- **taxpayers**
- **land registry documentation**



Population

1. **Link similar landmark assertions** from sources with a **root landmark**: use (pseudo-)identity criterion to link them
2. **Order landmark assertions**
3. **Infer changes and events** related to landmarks (Appearance, Disappearance)



Detection of similar landmark within sources

Population

4. [OPTIONAL] Inferring landmark identity (if not known at step 1)

Example: in the land registry, if a plot is split, its successors keep the same number in the registers

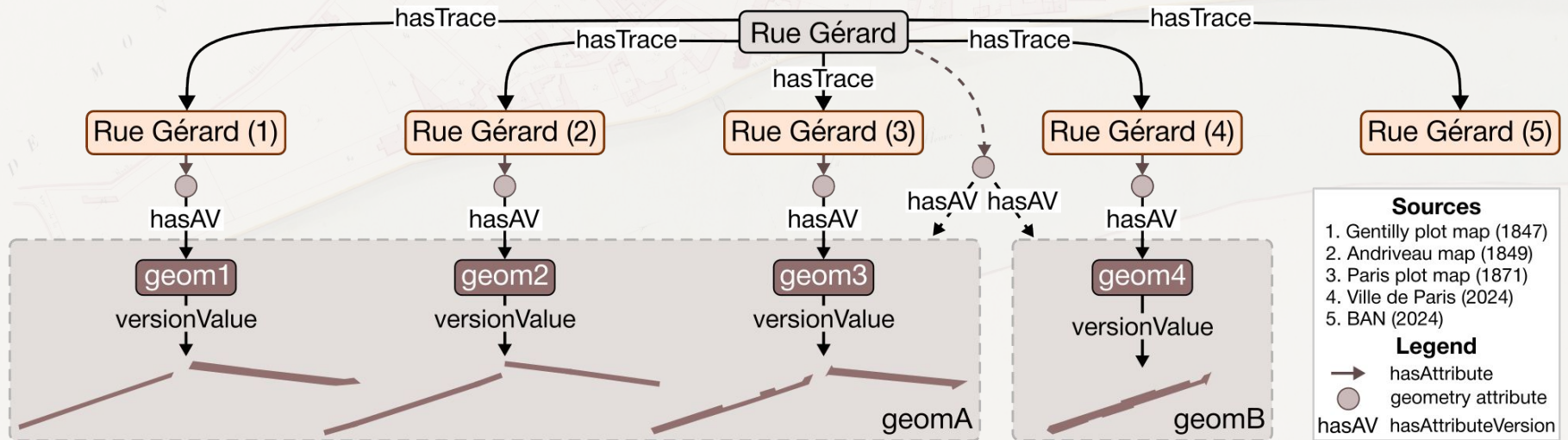


Population

5. Merge landmark attribute versions

6. Create changes and events that affect attributes of landmarks

The process can be repeated to infer new knowledge.

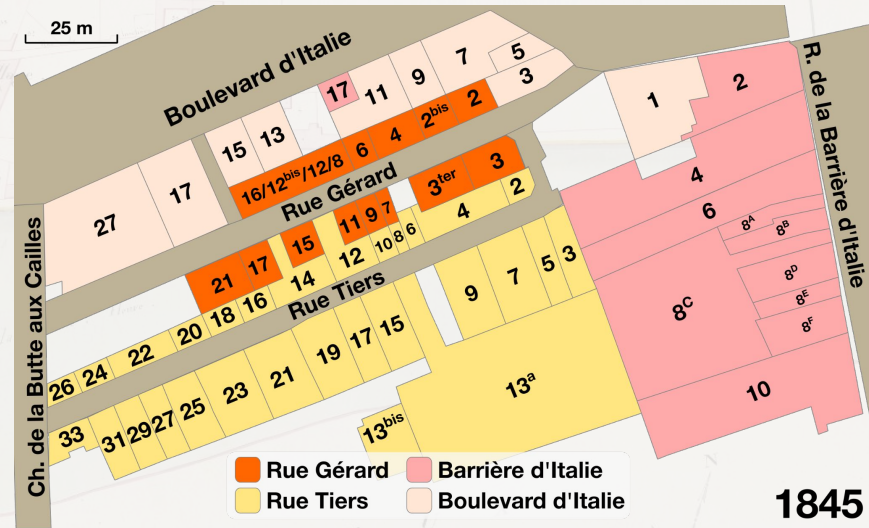


Example of merging attribute versions for Rue Gérard

Evaluation

- Answer **competency questions**
- Check **internal consistency** of data: SHACL rules and SPARQL queries
- Evaluate **information consistency**:
 - **difficult to assess quality** as no ground truth exists
 - ⇒ snapshots
 - ⇒ multi-source mapping (ex: addresses in several sources)

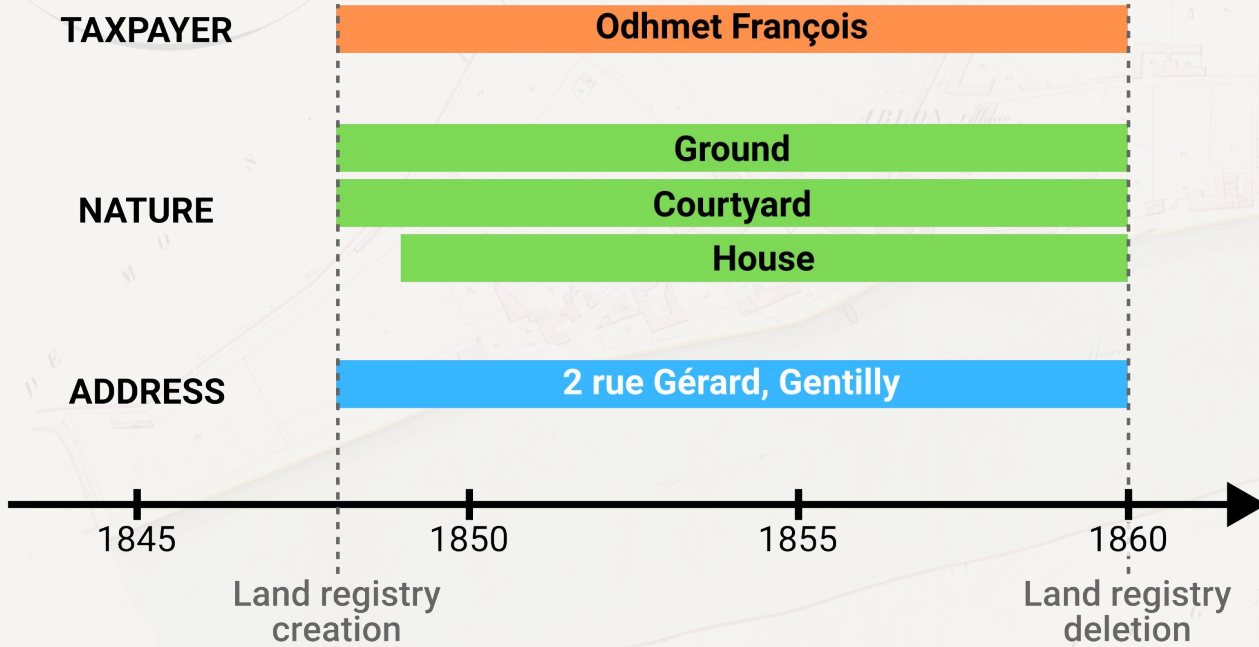
*Which streets already exist in 1845?
What were the plot addresses in 1845?*



Snapshot of the Butte aux Cailles neighborhood in 1845

Evaluation

Evolution of the attributes of a landmark



Plot D-14, Gentilly (second land registry)

What are the successive natures/addresses/taxpayers of plot D-14 in Gentilly?

Contributions

- **Ontology**

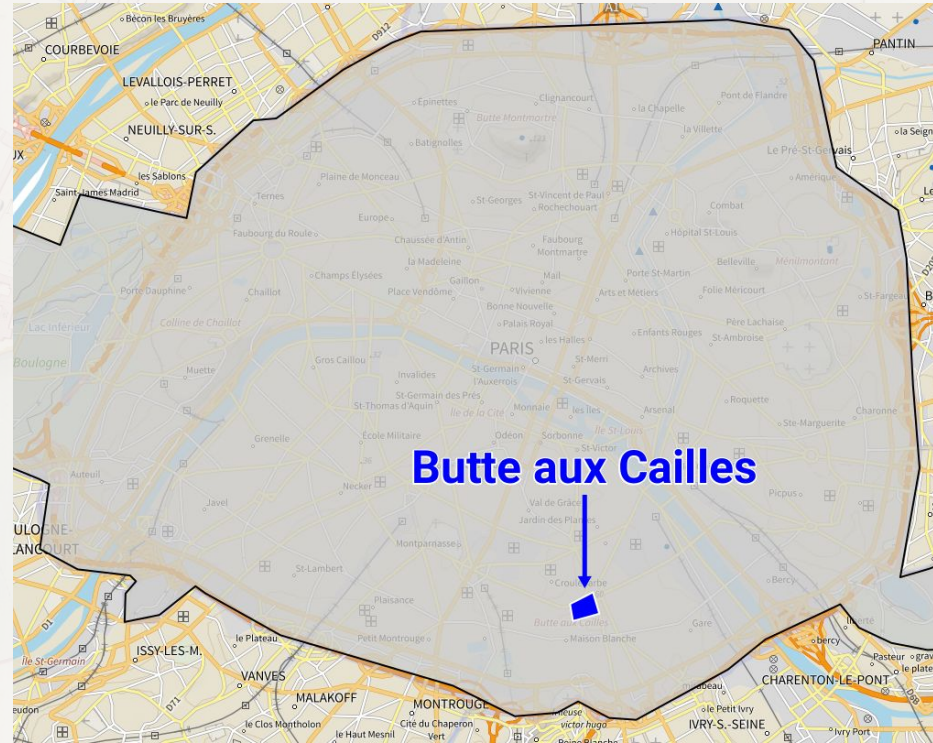
- describes addresses and landmarks
- **represents the evolution of their attributes**

- **Approach** to build a perpetual gazetteer as a KG

- treats several heterogeneous sources
- handles heterogeneous fragmentary data

Future work

- Use case on Butte aux Cailles neighborhood:
 - intend to **enlarge spatial cover** (whole Paris?)
 - add **new sources** (notary minutes)
- **Critical analysis** of the KG
 - detect inconsistencies
 - conflicts between sources



Location of the Buttes aux Cailles neighborhood within Paris

Future work

- Tools to **easily ask questions** to the graph: web map, natural language

Evolution des parcelles, voies et adresses de la Butte-aux-Cailles

Cadastrer

Cadastre de 1810 Cadastre de 1845

Informations sur les cadastres :

- Cadastre de 1810 : plan établi en 1810, matrice n°1 valide de 1813 à 1836; matrice n°2 valide de 1836 à 1848.
- Cadastre de 1845 : plan établi en 1845; matrice n°1 valide de 1848 à 1860.

Cliquez sur une parcelle pour en savoir plus.

Période

1813 1848

Le filtre temporel permet de visualiser la nature d'une parcelle au cours d'une année donnée.

Attribut

Choisir un attribut :

Agrégations associées à la parcelle sélectionnée

Voir les agrégations sous la forme suivante :

timeline liste

AGG_49685b6b-3d4

AGG_92fae1d1-fc8e-4de4-e

AGG_db6abec1-8337-4abf-b41b-f8b013d

A web map to query the KG (under development)

PeGaZus

- **Ontology**

- describes addresses and landmarks
- **represents the evolution of their attributes**

- **Approach** to build a perpetual gazetteer as a KG

- treats several heterogeneous sources
- handles heterogeneous fragmentary data

**Thank you for
your attention !**

Ontology, data and scripts are available

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

charly.bernard@ign.fr ; solenn.tual@ign.fr



PeGaZus

References

- [1] Kauppinen, T., Väättäin, J., Hyvönen, E. (2008). Creating and Using Geospatial Ontology Time Series in a Semantic Cultural Heritage Portal. In: Bechhofer, S., Hauswirth, M., Hoffmann, J., Koubarakis, M. (eds) The Semantic Web: Research and Applications. ESWC 2008. Lecture Notes in Computer Science, vol 5021. Springer, Berlin, Heidelberg.
https://doi.org/10.1007/978-3-540-68234-9_11
- [2] Bernard, Camille, Villanova-Oliver, Marlène, Gensel, Jérôme, Dao, Hy. : Modeling changes in territorial partitions over time: Ontologies tsn and tsn-change. In Proceedings of the 33rd Annual ACM Symposium on Applied Computing, pp. 866–875. Pau, France (2018).
- [3] Charles, William, Aussenac-Gilles, Nathalie, Hernandez, Nathalie. : HHT: an approach for representing temporally-evolving historical territories. In European Semantic Web Conference, pp. 419–435. Cham: Springer Nature Switzerland, (2023).
- [4] 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>

Ontology, data and scripts are available

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

charly.bernard@ign.fr ; solenn.tual@ign.fr

