

# Une approche pour la création d'un graphe spatio-temporel à partir des données extraites des annuaires: application aux photographes

Solenn Tual (1,2).

Encadrement : Nathalie Abadie (2), Joseph Chazalon (3), Bertrand Duménieu (4).

Séminaire SoDUCo - BnF, 10 novembre 2022



(1)



(2)



(3)



(4)



SODUCO ANR-18-CE38-0013



# Annuaires du commerce de Paris

## CONTENU

Une entrée peut contenir :

- Nom d'un individu, d'un commerce ou d'une organisation
- Description de l'activité (+/- complète)
- Adresse : voie, numéro de voie, type de lieux
- Titres (distinctions militaires, médailles professionnelles...)

Cherot, joaillier. S. Martin, 51.

Chéroux, orf. Sté Avoye, 42.

Cherre, layet. Caire, 7.

Cherré, tap. Moulins, 42.

Cherrier, tabl. St Denis, 277.

Chevalier, opt. du Roi, Tour de l'Horl.  
du Palais, 1.

Chevalier (L.), opt. q. Horl. Pal. 65.

Chevalier (Vinc.), opt. q. Horl. Pal. 69.

Chevalier (Victor), opt. q. Horl. Pal. 77 b.

**Liste ordonnée par nom - Annuaire Cambon Almgène 1839**

## Annuaires du commerce de Paris

### STRUCTURE

- Entrées structurées dans différentes listes (classées par nom, par profession, par adresse)
- Redondance d'une année ou d'une édition à une autre

Bibliothèque Ste-Genèviève, Clotilde, 1.

Bibliothèque de la Ville, quai d'Austerlitz, 33  
(provisoirement).

Biblique protestante (Société), Moulins, 16.

Bibron, aide-natural., au Muséum d'hist. nat.

Bibus, tailleur, Roule, 21.

Bibus, tailleur, Richelieu, 31.

Bical et Dorre, fab. de socques, Vertbois, 14.

Bicant (Mme), fondeur en cuivre, cour de la  
Corderie-du-Temple, 26.

Bichard (Mme), Nve-de-Luxembourg, 17.

Bichard, tabacs et eau-de-vie, Faub.-St-Martin, 45.

Bibliothèque Ste-Genèviève, Clotilde, 1.

Bibliothèque de la Ville, quai d'Austerlitz, 33  
(provisoirement).

Biblique protestante (Société), Moulins, 16.

Bibron, aide-natural., au Muséum d'hist. nat.

Bibus, tailleur, Roule, 21.

Bibus, tailleur, Richelieu, 31.

Bicai, fab. de jouets, Montmorency, 33.

Bical et Dorre, fab. de socques, Vertbois, 14.

Bican (Vve) et fils, fondeur en cuivre, place  
de la Corderie-du-Temple, 26.

Bicel, épicier, marché d'Aguesseau, 15.

Richard (Mme), Nve-de-Luxembourg, 17

Bichard, tabacs et eau-de-vie, Faub.-St-Martin, 45.

Bibliothèque Ste-Genèviève, rue des Sept-Voies  
et place du Panthéon.

Bibliothèque de la Ville, quai d'Austerlitz, 33  
(provisoirement).

Bibus, tailleur, Richelieu, 31.

Bical, fab. de jouets, Montmorency, 33.

Bical et Doire, fab. de socques, Vert-Bois, 14.

Bican (Vve) et fils, fondeurs en cuivre, place  
de la Corderie-du-Temple, 26.

Bicel, épicier, Marché-d'Aguesseau, 15.

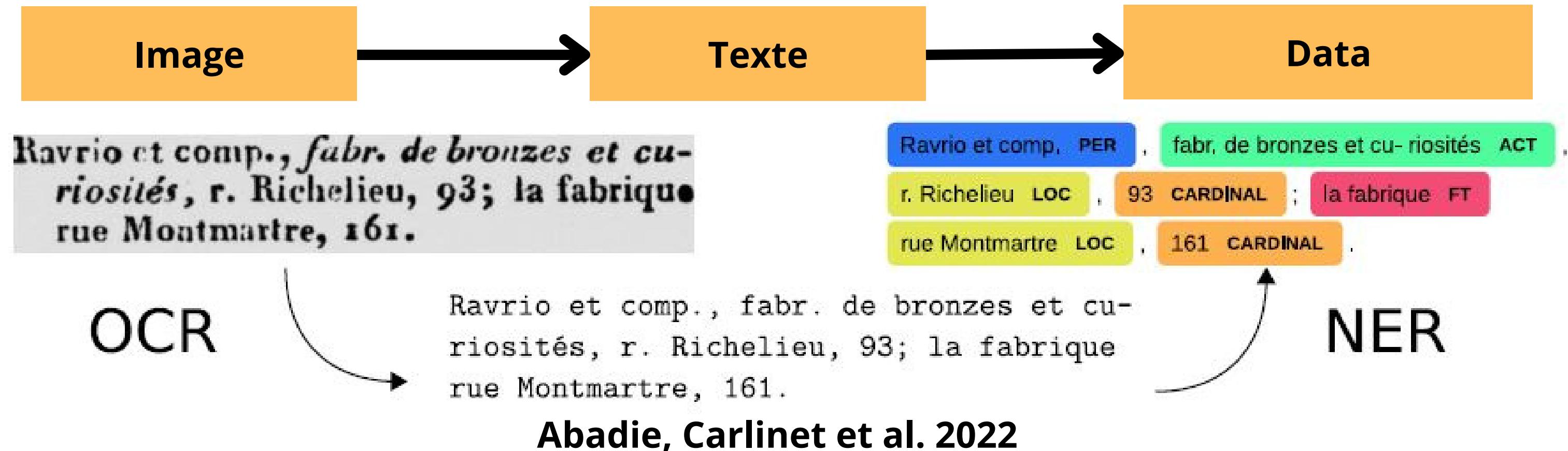
Bichard, tabac et eau-de-vie, Faub.-St-Martin, 45.

Didot 1841a - page 95

Didot 1842a - page 117

Didot 1843a - page 129

# Approche d'extraction pré-existante



OCR : Reconnaissance Optique de Caractères

→ Transcription du texte contenu dans les images en texte éditable

NER : Reconnaissance des Entités Nommées

→ Recherche et classification de mots ou groupes de mots relatifs à certains types d'informations

## Annuaires du commerce de Paris

**DONNEES**

Données extraites avec la chaîne de traitement :  
9 821 898 entrées

- Structurées dans une base de donnée relationnelle
- Requêtables (SQL)

# Annuaires du commerce de Paris

## DONNEES

Données extraites avec la chaîne de traitement :  
9 821 898 entrées

- Structurées dans une base de donnée relationnelle
- Requêtables (SQL)

## OBJECTIF

Exploiter les données extraites :

- Redondance temporelle
  - Informations spatiales
- Graphe spatio-temporel

## Point de départ

Extraction des entrées :  
Segmentation + OCR + NER => Données structurées

# Objectifs et démarche

## Point de départ

Extraction des entrées :  
Segmentation + OCR + NER => Données structurées

**Cas d'application : photographes  
et professions associées  
(fournisseurs)**

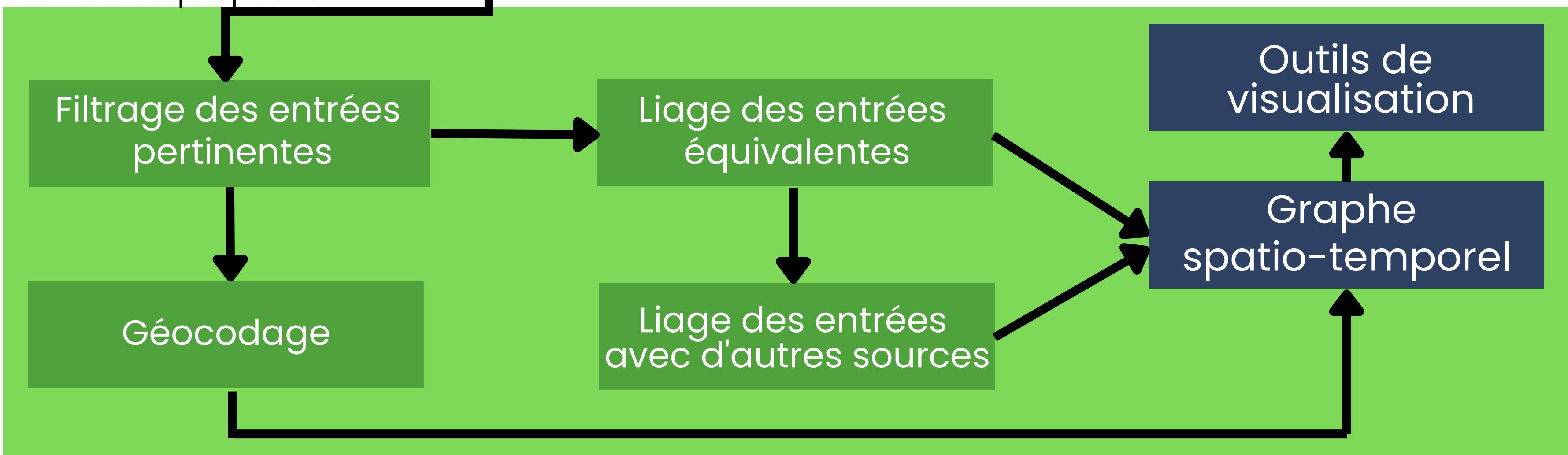
# Démarche

Point de départ

Extraction des entrées :  
Segmentation + OCR + NER => Données structurées

Cas d'application : photographes  
et professions associées  
(fournisseurs)

Démarche proposée

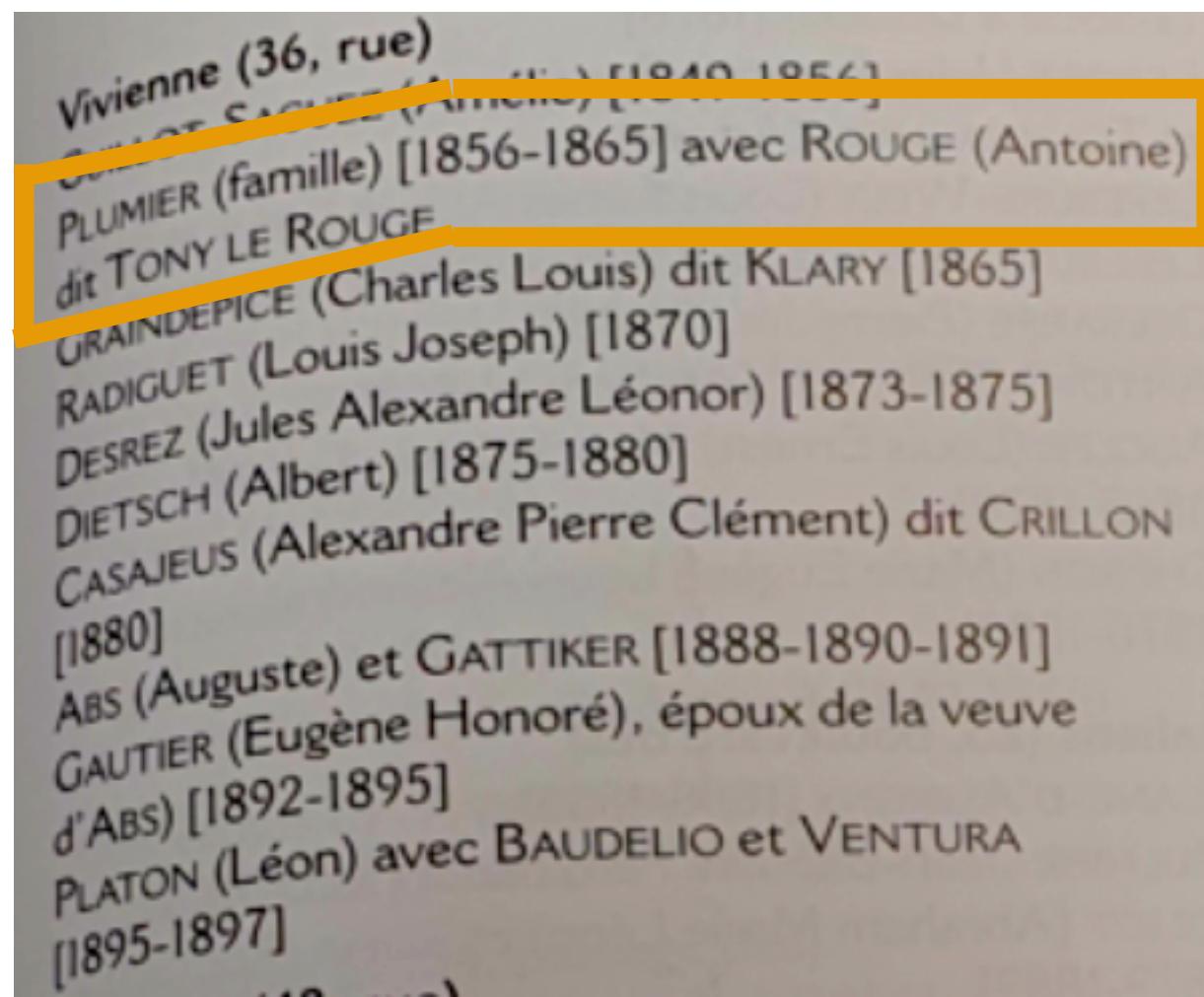


## Contexte

- I Sélection des données pertinentes
  - II Création du graphe spatio-temporel
  - III Visualisation des données
- Conclusion

## I. Sélection des données pertinentes

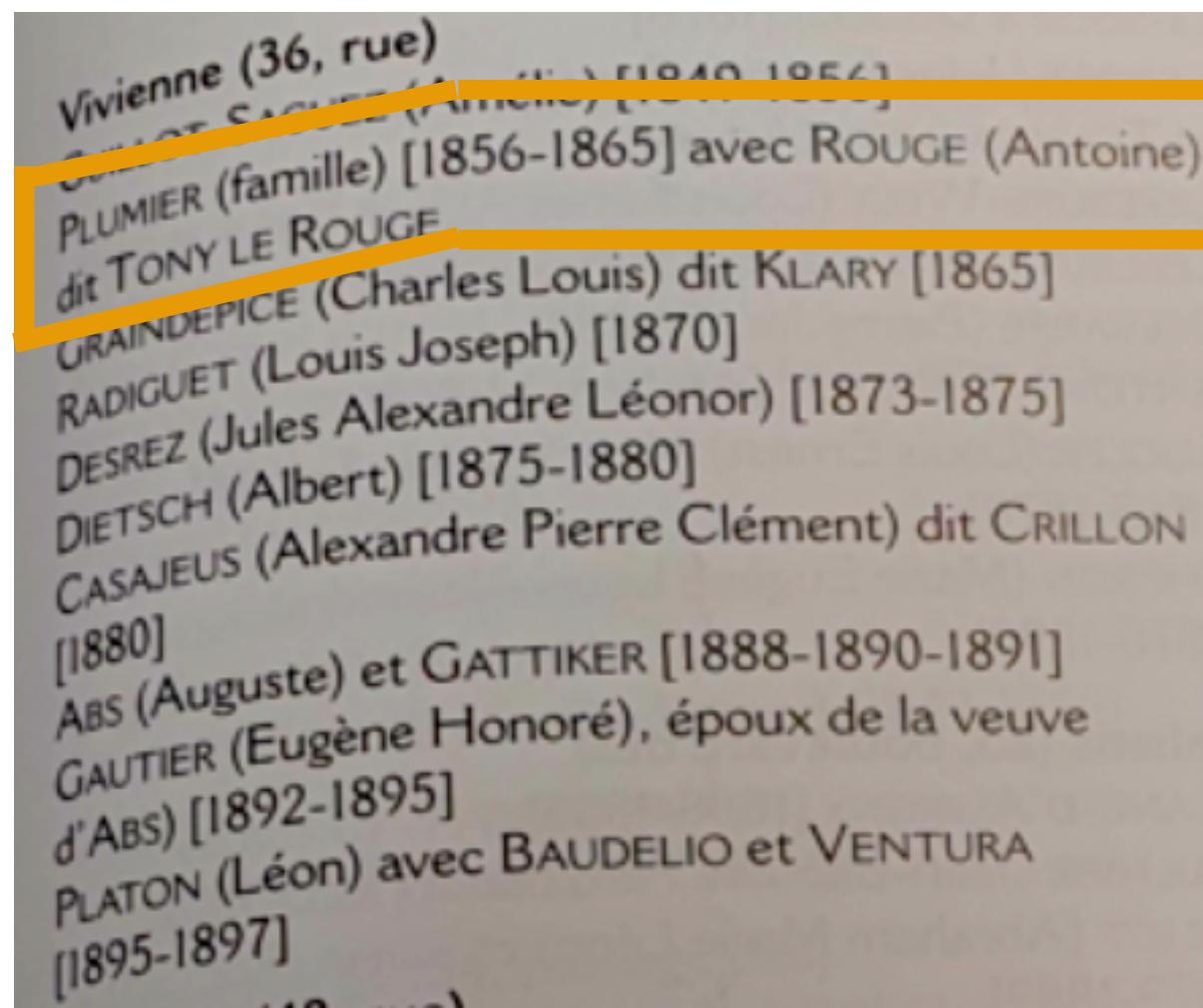
- Constitution d'une liste de mots-clés utilisés pour filtrer les entrées pertinentes pour notre cas d'étude
- Recherche des photographes et ateliers (~230) listés par des historiens de l'art (Durand, 2015) dans la base de donnée des extractions



Marc Durand (dir.), « De l'image fixe à l'image animée: 1820-1910. Tome 2: actes des notaires de Paris pour servir à l'histoire des photographes et de la photographie ». Archives nationales (2015).  
Pierrefitte-sur-Seine.

## I. Sélection des données pertinentes

- Constitution d'une liste de mots-clés utilisés pour filtrer les entrées pertinentes pour notre cas d'étude
- Recherche des photographes et ateliers (~230) listés par des historiens de l'art (Durand, 2015) dans la base de donnée des extractions



Vivienne (36, rue)  
GUYOT SAGNER (Amélie) [1849-1856]  
PLUMIER (famille) [1856-1865] avec ROUGE (Antoine)  
dit TONY LE ROUGE  
URAINDEPICE (Charles Louis) dit KLARY [1865]  
RADIGUET (Louis Joseph) [1870]  
DESREZ (Jules Alexandre Léonor) [1873-1875]  
DIETSCH (Albert) [1875-1880]  
CASAJEUS (Alexandre Pierre Clément) dit CRILLON  
[1880]  
ABS (Auguste) et GATTIKER [1888-1890-1891]  
GAUTIER (Eugène Honoré), époux de la veuve  
d'ABS) [1892-1895]  
PLATON (Léon) avec BAUDELIO et VENTURA  
[1895-1897]



```
SELECT *  
FROM directories.elements AS e  
WHERE e.persons ILIKE '%plumi%'  
ORDER BY e.published
```

*Exemple de requête SQL  
Toutes les entrées dont l'entité  
"Nom de personne/commerce"  
contient "plumi"*

## I. Sélection des données pertinentes

- Constitution d'une liste de mots-clés utilisés pour filtrer les entrées pertinentes pour notre cas d'étude

Recherche des photographes et ateliers (~230) listés par des historiens de l'art (Durand, 2015) dans la base de donnée des extractions

Vivienne (36, rue)
CHAPOT-SAGNER (Amélie) [1849-1856]
<b>PLUMIER (famille) [1856-1865] avec ROUGE (Antoine)</b>
dit TONY LE ROUGE
URAINDEPICE (Charles Louis) dit KLARY [1865]
RADIGUET (Louis Joseph) [1870]
DESREZ (Jules Alexandre Léonor) [1873-1875]
DIETSCH (Albert) [1875-1880]
CASAJEUS (Alexandre Pierre Clément) dit CRILLON [1880]
ABS (Auguste) et GATTIKER [1888-1890-1891]
GAUTIER (Eugène Honoré), époux de la veuve d'ABS) [1892-1895]
PLATON (Léon) avec BAUDELIO et VENTURA [1895-1897]



```
SELECT *
FROM directories.elements AS e
WHERE e.persons ILIKE '%plumi%'
ORDER BY e.published
```

*Exemple de requête SQL  
Toutes les entrées dont l'entité "Nom de personne/commerce" contient "plumi"*

Plumier (Victor), **portraits** sur plaques et sur pap., Vivienne, 36.

**Didot\_1856a**

*Exemple de résultat*

## I. Sélection des données pertinentes

- Constitution d'une liste de mots-clés utilisés pour filtrer les entrées pertinentes pour notre cas d'étude

 Liste de 19 mots-clés => 227 323 entrées.

Mots et expressions vissés	Mot-clé retenu	Mots et expressions vissés	Mot-clé retenu
Photographe, photographie, ...	photo	Instrument/outil/appareil de mathématiques, mathématicien	math
Daguéréotype	daguer	Physique	physique
Opticien, optique	opti	Produits chimiques	chimi
Lentille	lentil	Ingénieur	ingenieur
Lunettes, lunettier	lunet	Image	imag
Graveur	grav	Portrait	portr
Lithographe	litho	Prisme	prisme
Artiste	artiste	Appareil/instrument de mesure	mesure
Chambre noire	chamb AND noir	Camera obscura	camera
Cinéma, cinématographe	cinem		

## I. Sélection des données pertinentes

- Constitution d'une liste de mots-clés utilisés pour filtrer les entrées pertinentes pour notre cas d'étude

Réduction de la liste de mots-clés utilisés aux trois mots les plus couramment associés aux photographes listés dans la référence.

**PHOTO**

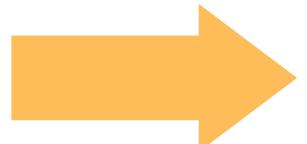
Photographe  
Photographie

**DAGUER**

Daguerréotype

**OPTI**

Optique  
Opticien



**34 062 entrées**

## Contexte

- I Sélection des données pertinentes
  - II Création du graphe spatio-temporel
  - III Visualisation des données
- ## Conclusion

Liage des entrées relatives aux mêmes personnes ou commerces

### Méthode logique

Appariement des entrées par **comparaison stricte** de clés

- Clé = combinaison de propriétés
  - Numéro de l'entrée (si plusieurs activités ou adresses dans l'entrée)
  - Nom et Activité
  - Nom et Adresse
  - Adresse et Activité
  
- Crédit à l'aide d'un raisonneur dans un triplestore\*

\*base de données "graphe" - contient uniquement des triplets RDF

### Liage des entrées relatives aux mêmes personnes ou commerces

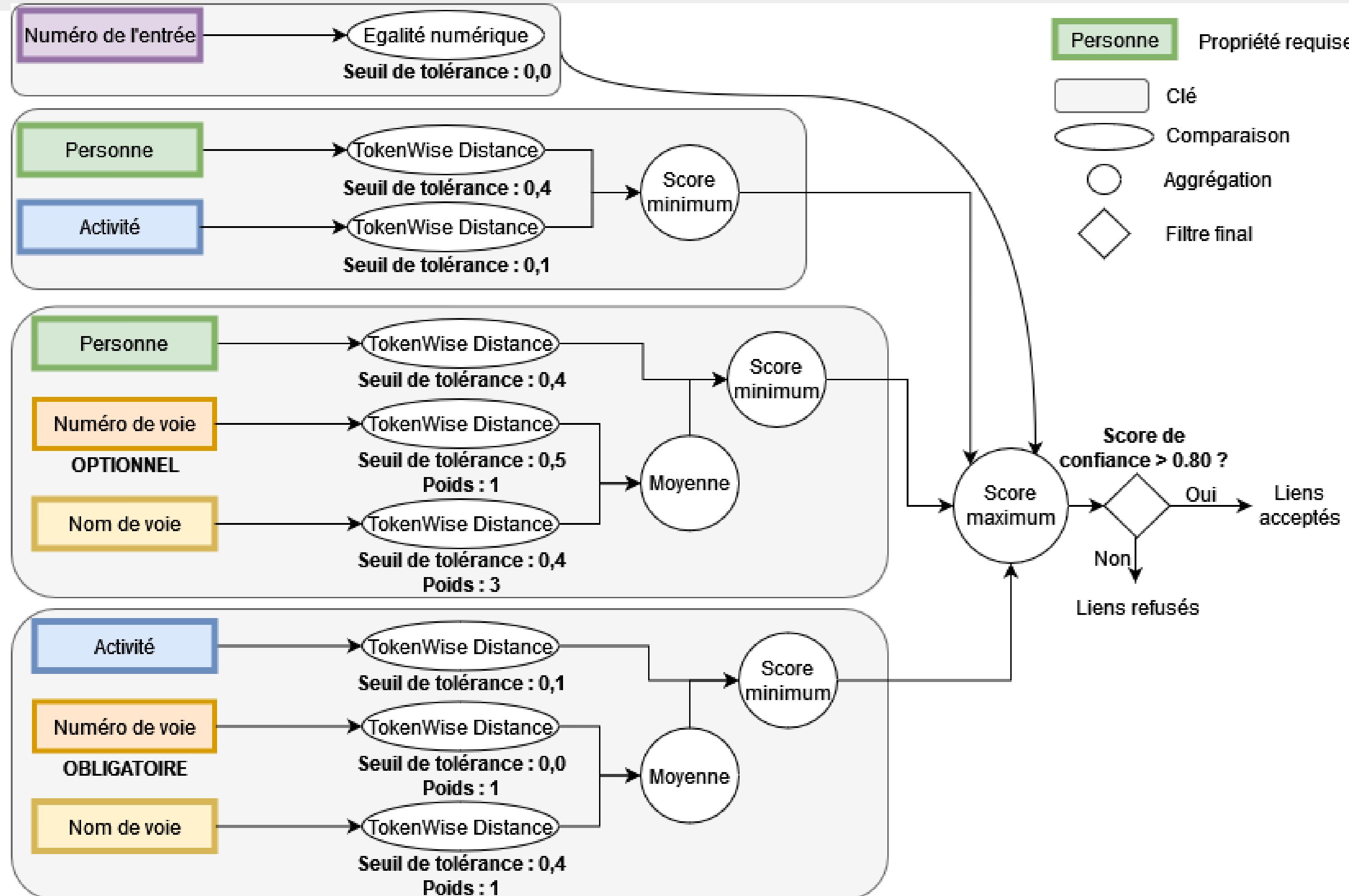
#### Méthode numérique

- Appariement par **comparaison numérique** des propriétés
- Distance d'édition TokenWise
  - Calcul de la distance de Levenshtein\* entre tous les mots du texte
  - Score final normalisé (entre 0 et 1)

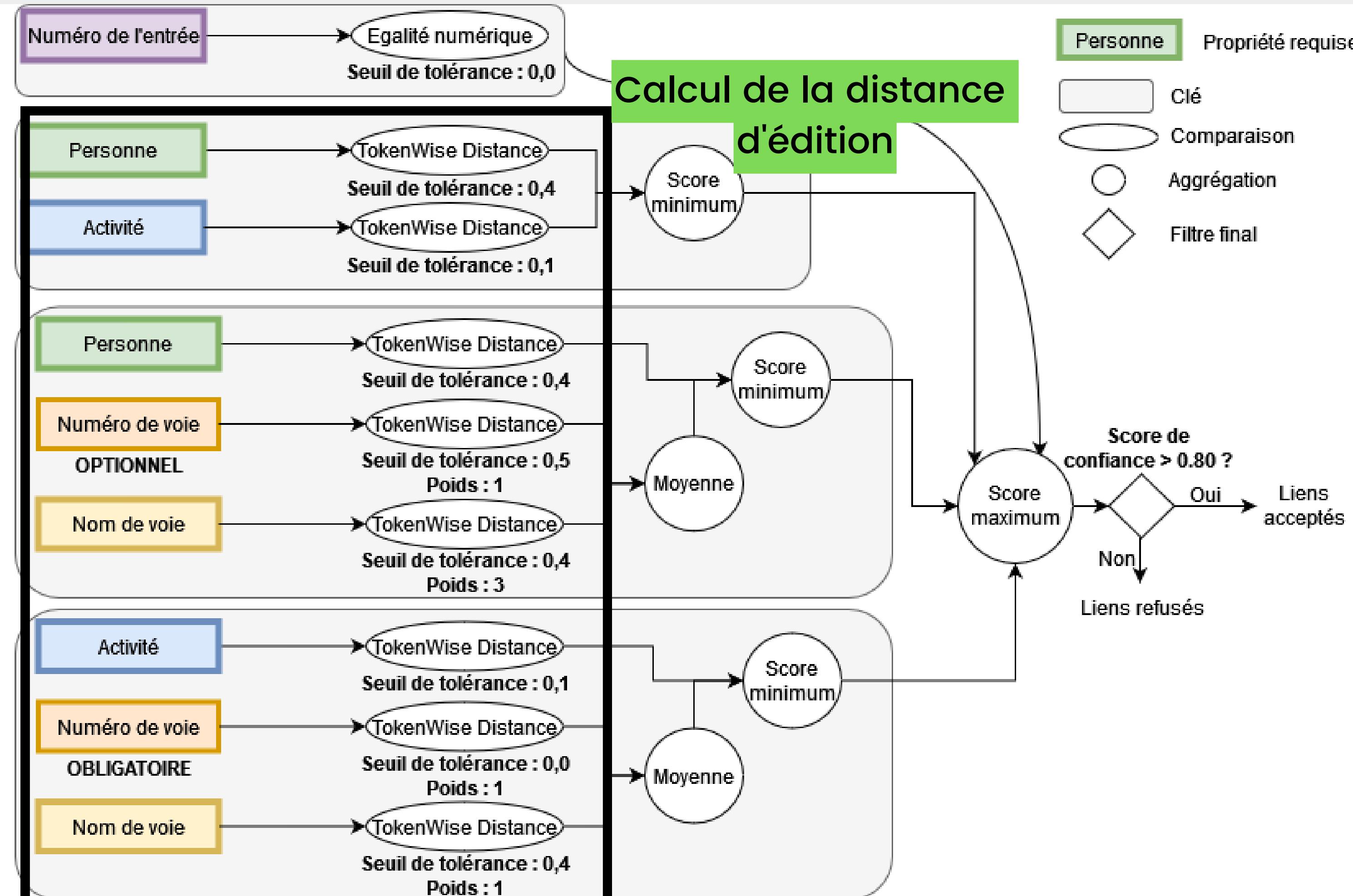
\* Distance de Levenshtein : compte le nombre d'inspections, de remplacements et de suppressions de caractères pour passer d'un mot à un autre

ex :    **CHAT** <=> **PLAT**              2 remplacements

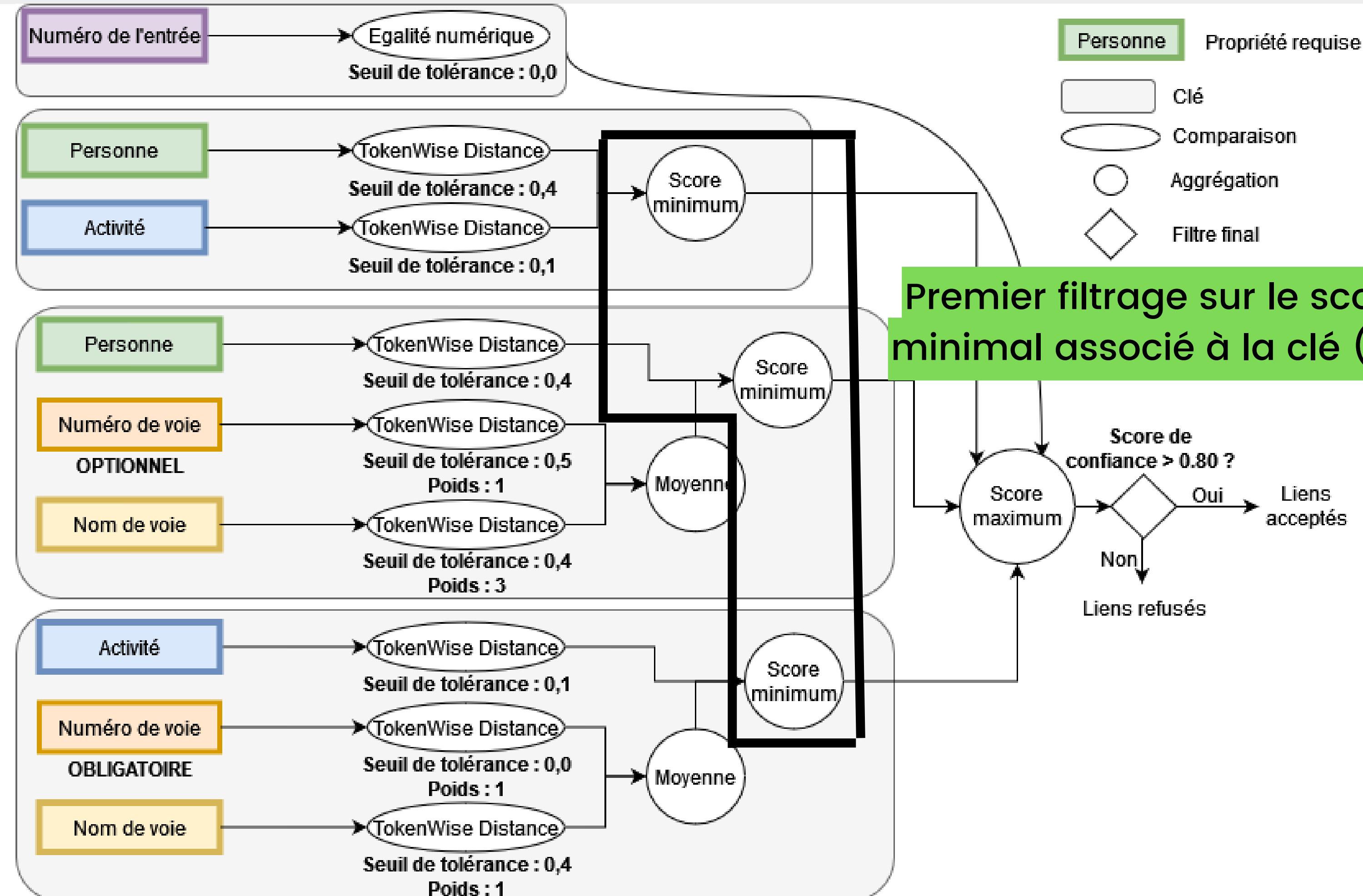
## II. Création du graphe spatio-temporel



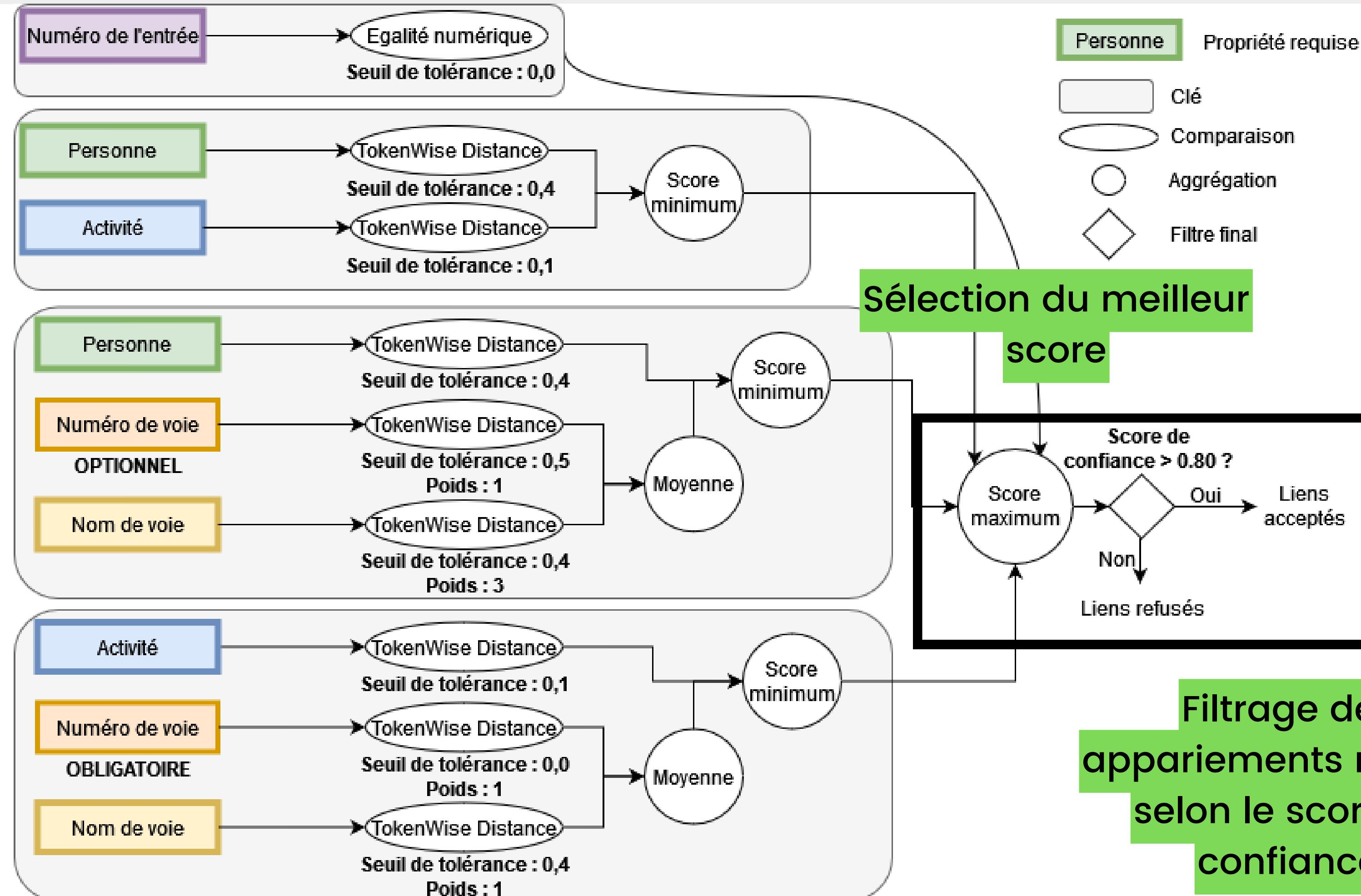
## II. Création du graphe spatio-temporel



## II. Création du graphe spatio-temporel



## II. Création du graphe spatio-temporel



## II. Création du graphe spatio-temporel

~ 38 000 ressources

### Méthode logique

**250 622 liens**

Paramétrage simple = clés

Comparaison très stricte

### Méthode numérique

& Propagation des liens sameAs

**357 130 liens**

Paramétrage complexe :  
identifier les seuils de tolérance pertinents

Comparaison plus adaptée aux  
chaînes de caractères résultant de  
l'OCR

### BILAN

**401 852 liens d'équivalence distincts entre les ressources du graphe**

→ Graphe spatio-temporel : Résultats des appariements +  
résultat du géocodage des entrées

### Liage avec d'autres ressources

→ Appariement par **comparaison numérique** des propriétés relatives au nom des photographes entre :



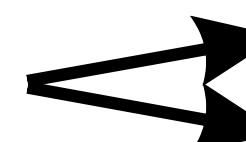
Annuaires



Raison sociale



DATA BNF



Label préféré

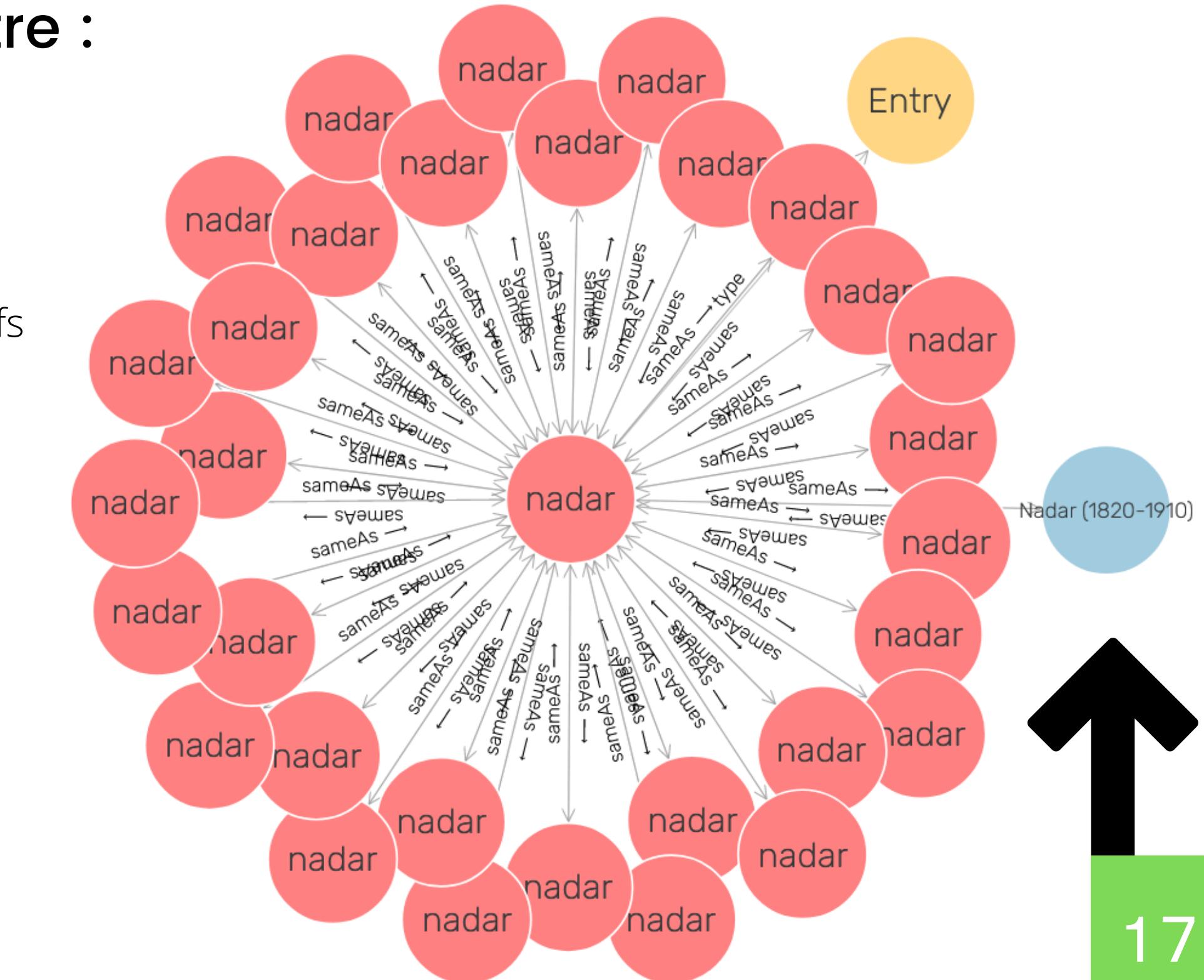
Labels alternatifs



Nadar (1820-1910) - Auteur - Ressources de la Bibliothèque nationale de France

Toutes les informations de la Bibliothèque Nationale de France sur :  
Nadar (1820-1910)

[data.bnf.fr](#)



## Contexte et objectifs

- I Sélection des données pertinentes
  - II Constitution du graphe-spatio-temporel
  - III Visualisation des résultats
- Conclusion

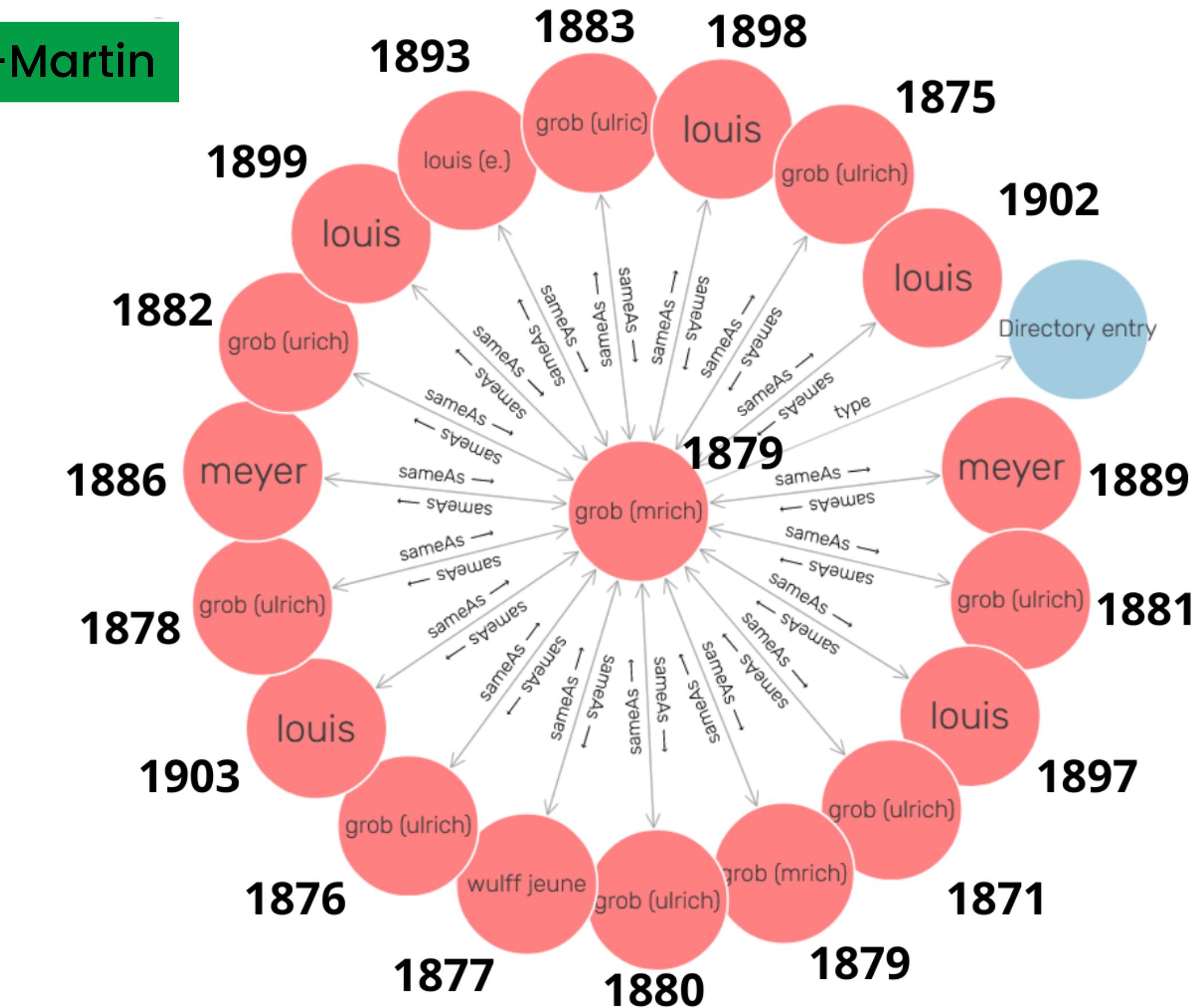
### III. Visualisation des résultats

#### Graphe spatio-temporel

Succession : 29 boulevard Saint-Martin

Méthode logique

19 ressources



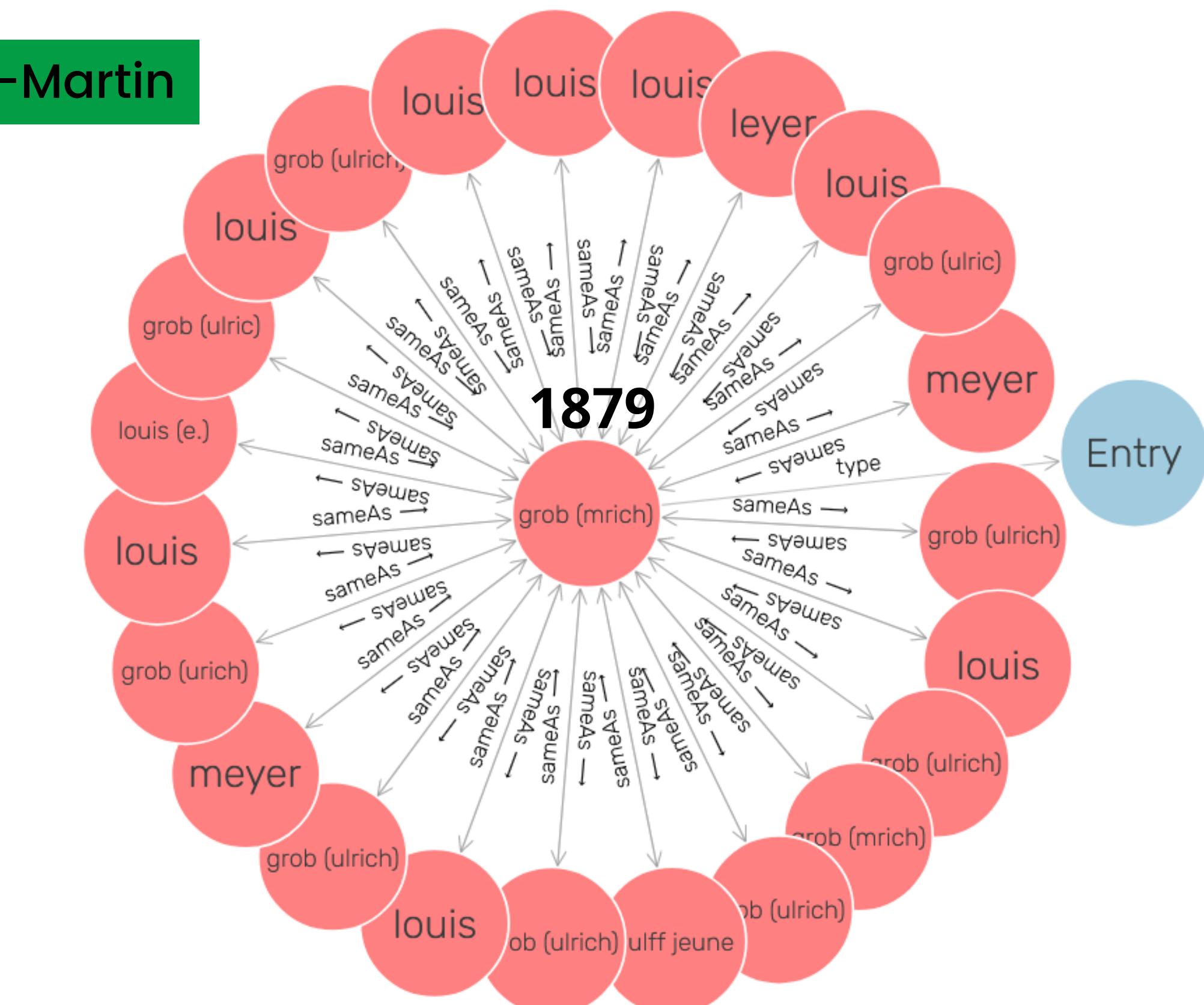
### III. Visualisation des résultats

## Graphe spatio-temporel

Succession : 29 boulevard Saint-Martin

Méthode logique +  
Méthode numérique

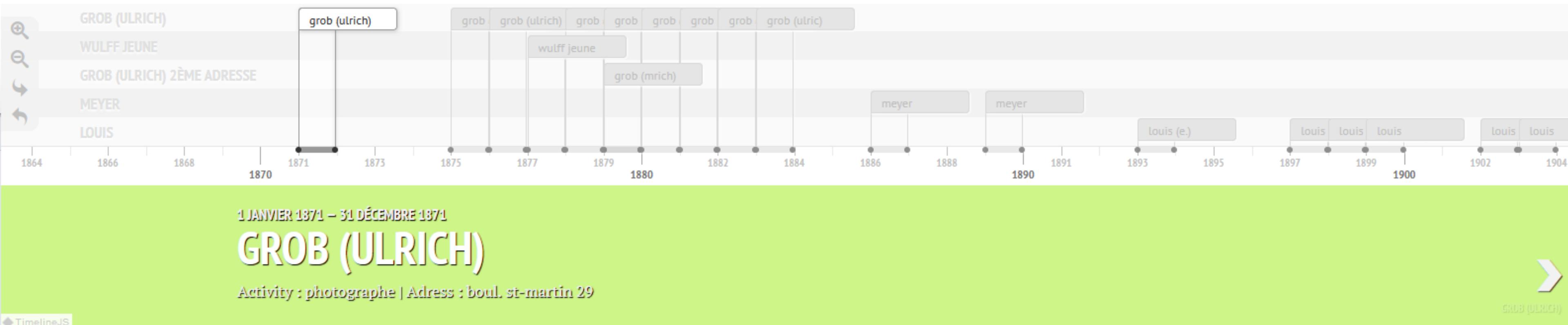
24 ressources



### III. Visualisation des résultats

## Graphe spatio-temporel

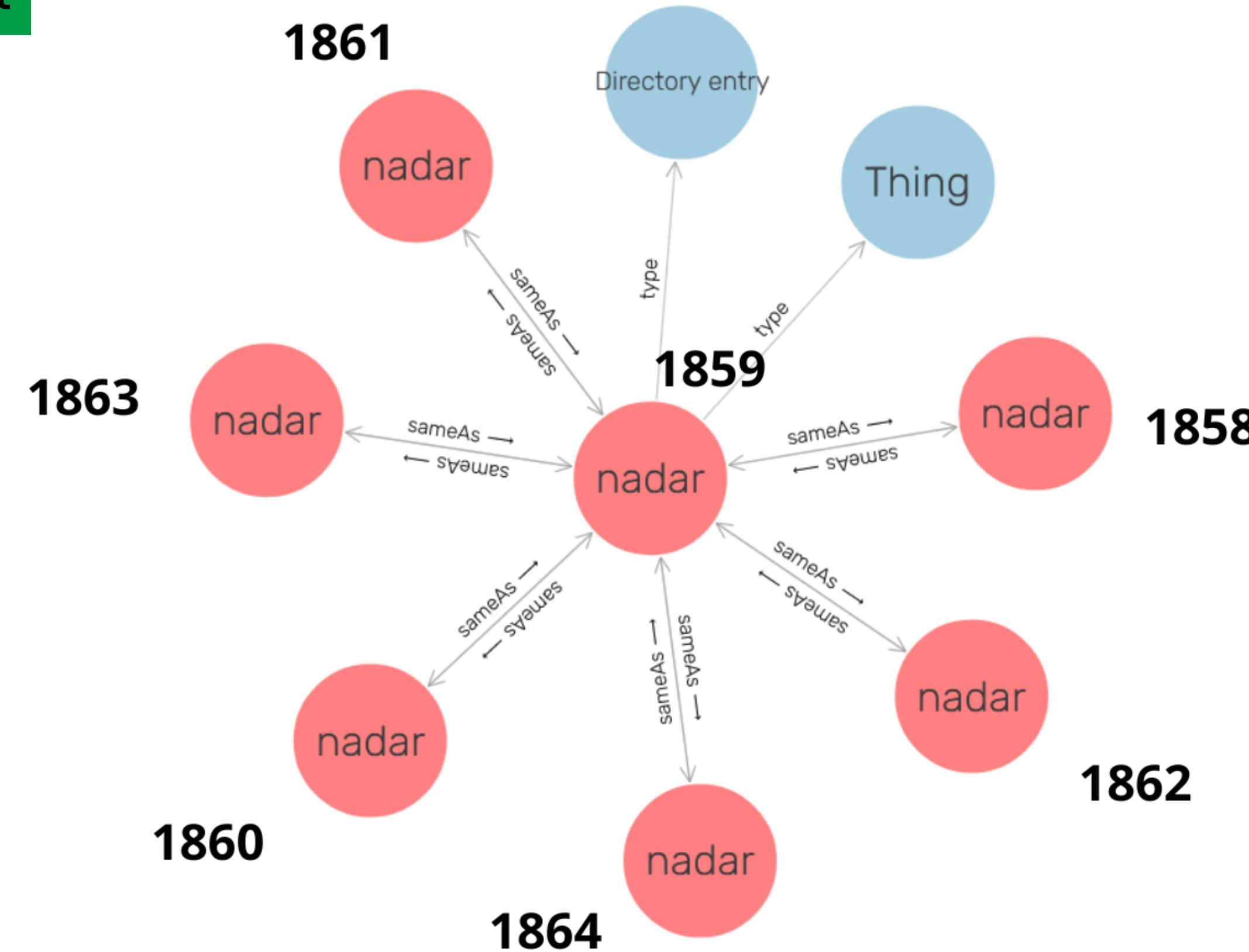
Succession : 29 boulevard Saint-Martin



### III. Visualisation des résultats

## Graphe spatio-temporel

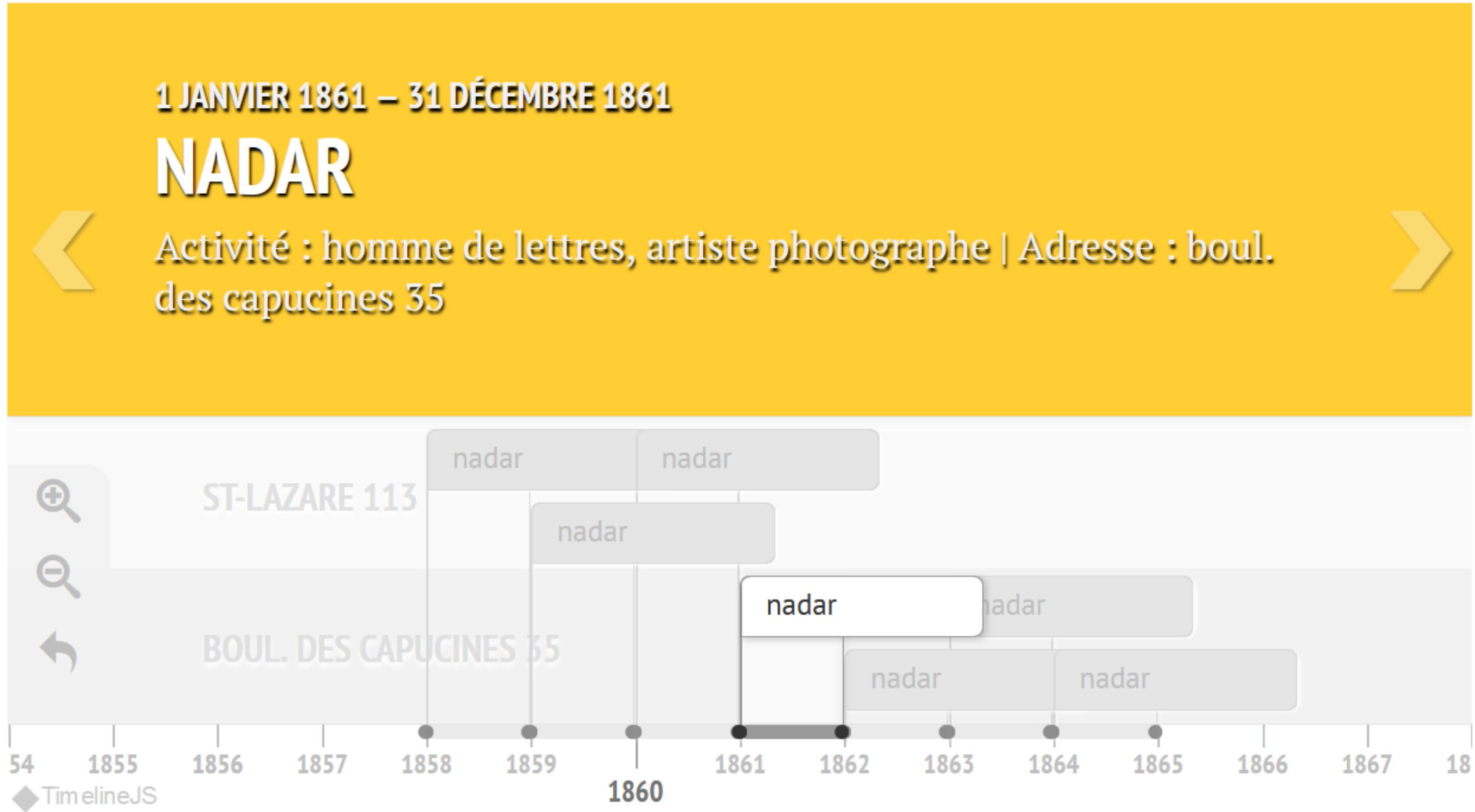
### Déménagement



### III. Visualisation des résultats

## Graphe spatio-temporel

### Déménagement



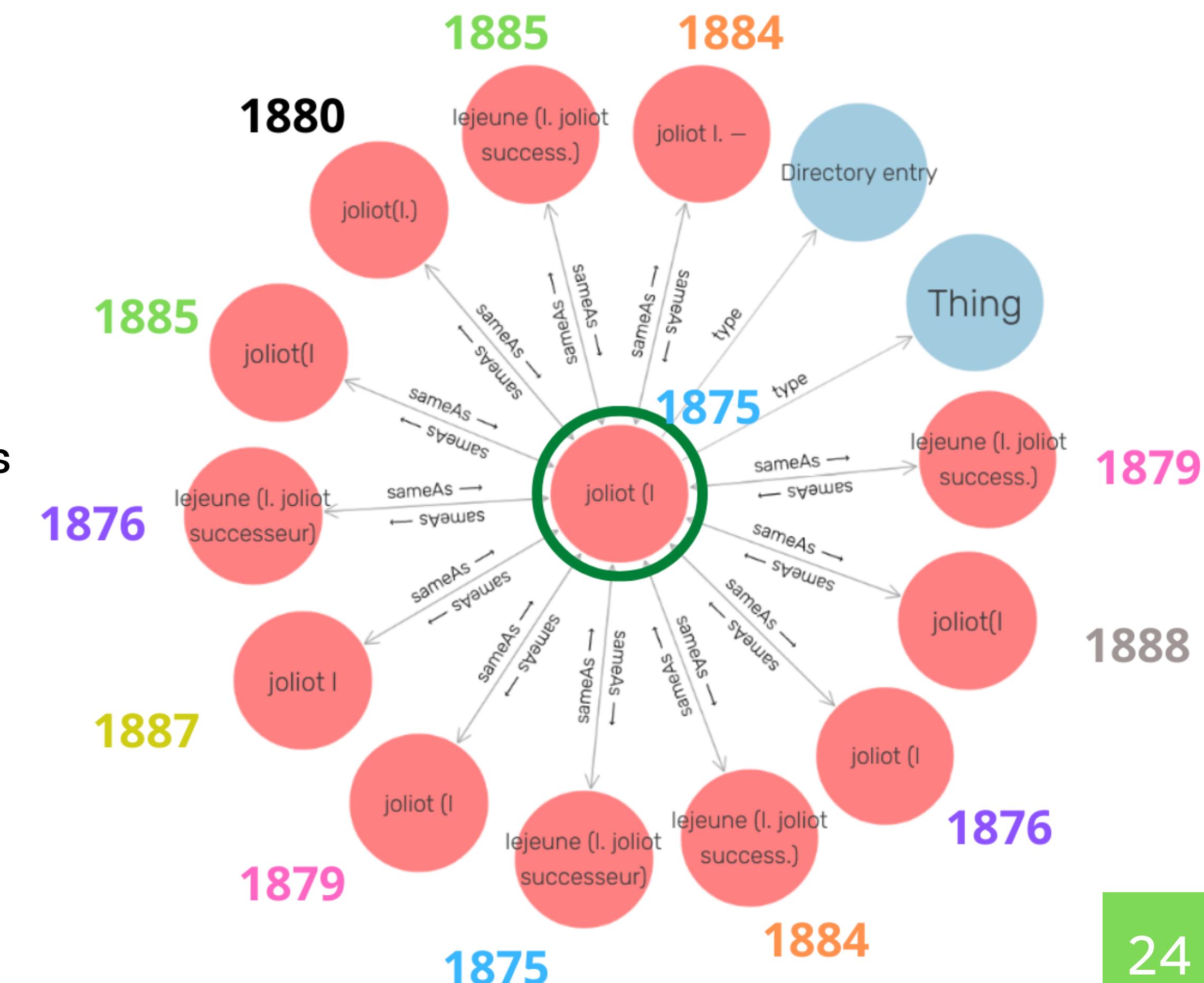
### III. Visualisation des résultats

## Graphe spatio-temporel

### Succession

Entrées relatives au photographe  
Lejeune et à son successeur

- Appariements entre listes
- Appariements entre annuaires



### III. Visualisation des résultats

## Visualisation cartographique

DEMO

### Recherche avancée

A historical map of Paris, specifically the 10th arrondissement, showing street grids and building footprints. A search result for 'grob (mrich)' is displayed as a callout box. The box contains the following information:

**grob (mrich)**

Adresse (annuaire) : boul. st-martin 29  
Adresse (géocodage) : 29 boulevard saint martin  
Activité : photographe  
Année de publication : 1879  
Annuaire : DidotBottin\_1879  
Identifiant de l'entrée : 4080943

The map also features a zoom control on the top left, a scale bar at the bottom left (500 m / 2000 ft), and a Leaflet logo at the bottom right.

#### Filtres

##### Paramètres de recherche

Laissez les champs vides puis cliquez sur "Valider" pour faire une recherche sur l'intégralité du jeu de données.

##### Raison sociale

grob

##### Activité

Mot-clé

##### Adresse

Adresse

Valider

##### Filtrer par période

1879

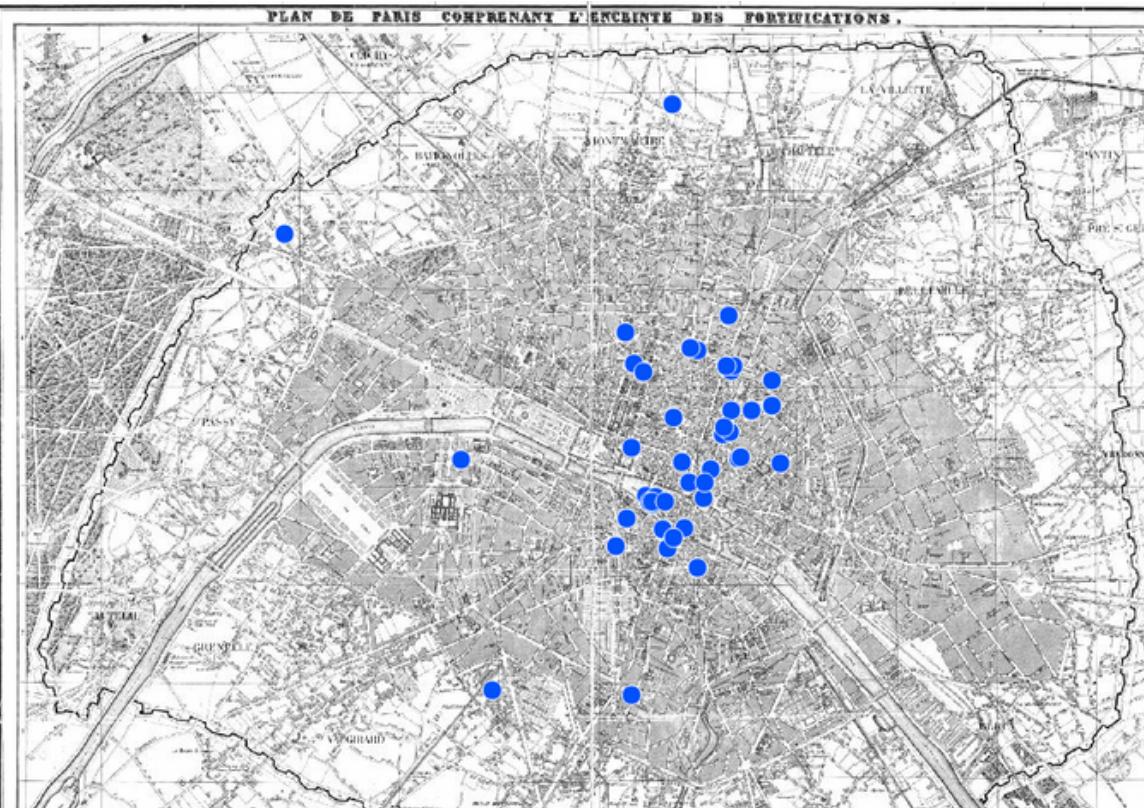
1879



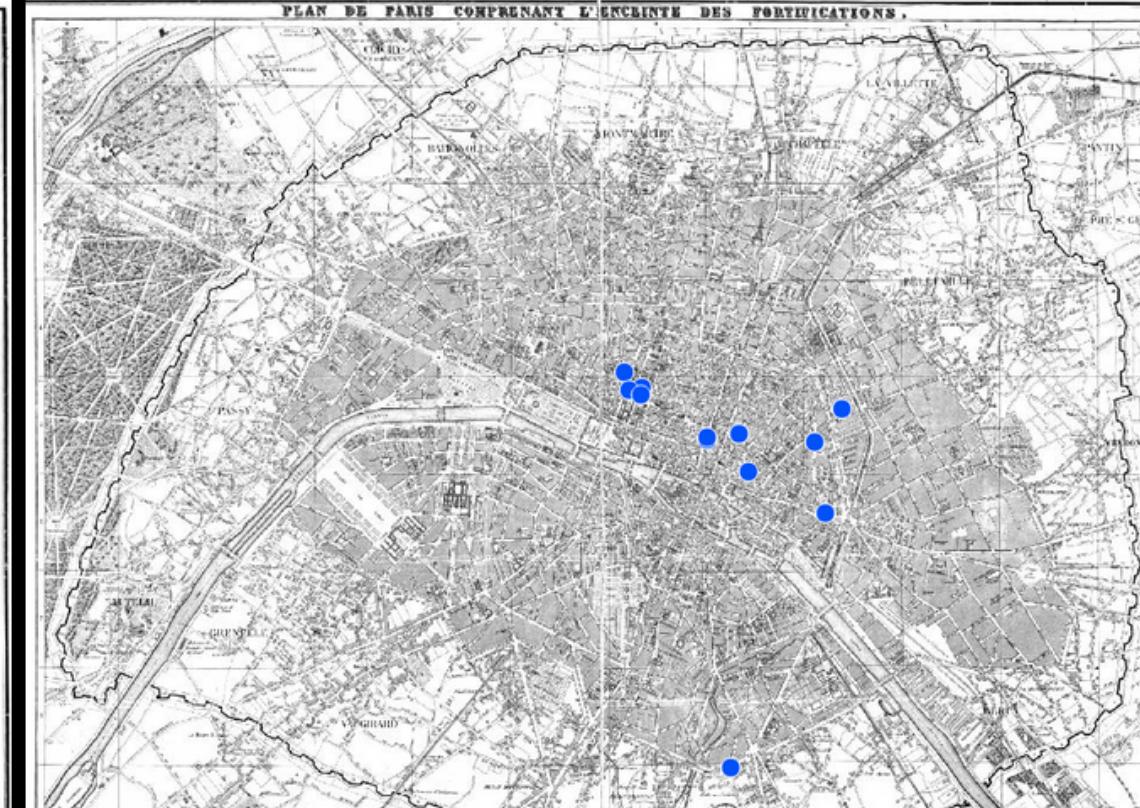
### III. Visualisation des résultats

## Visualisation cartographique

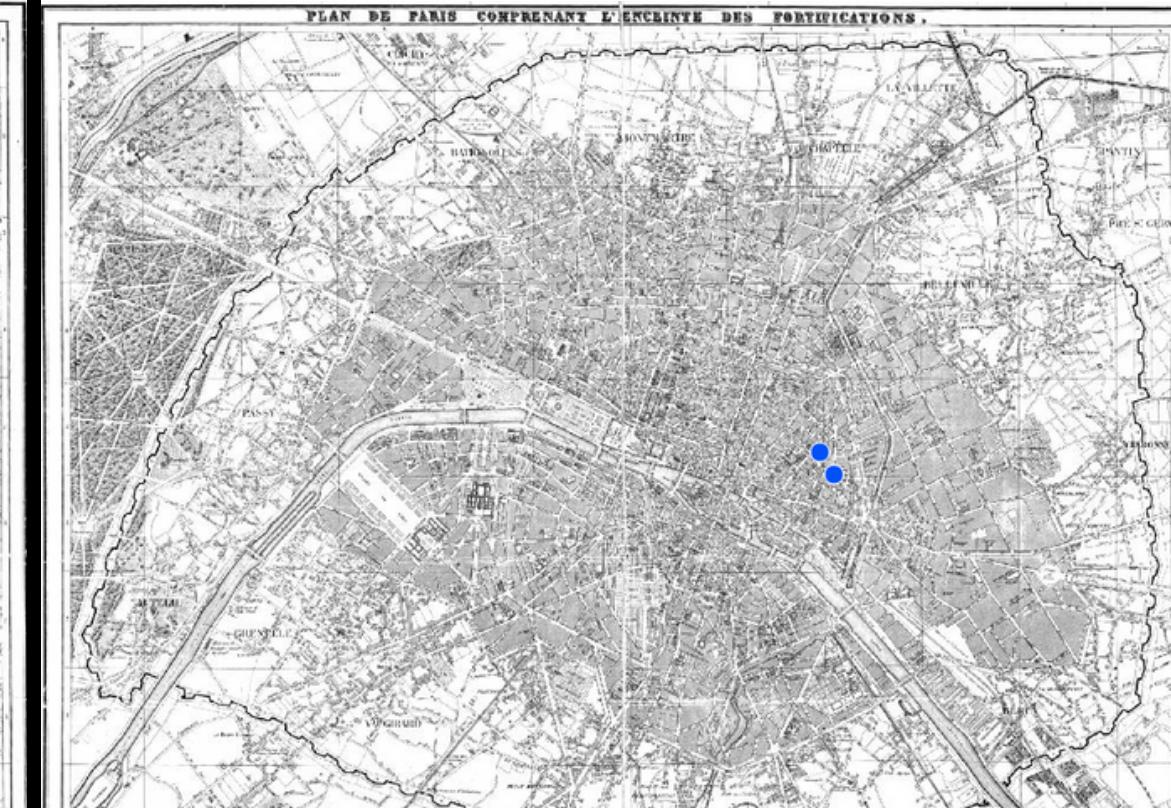
1845



1860

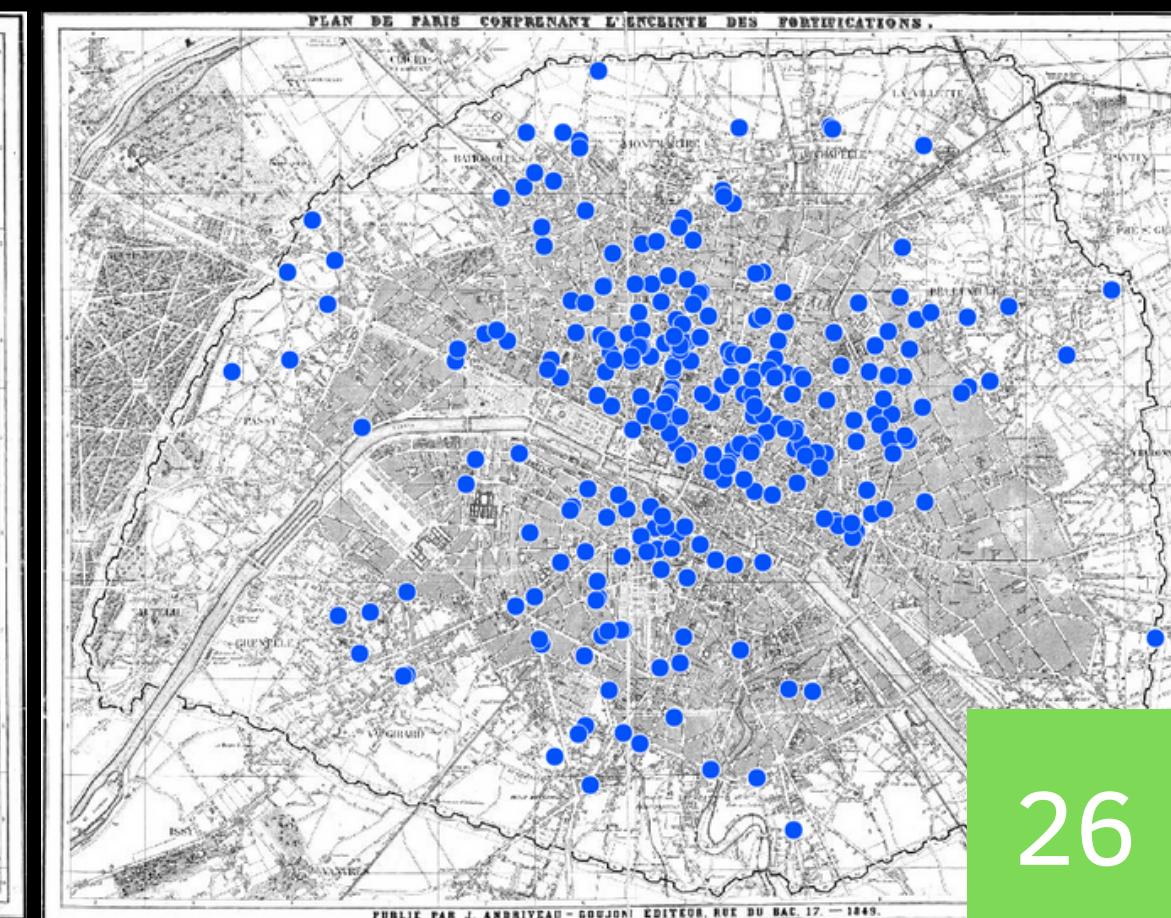
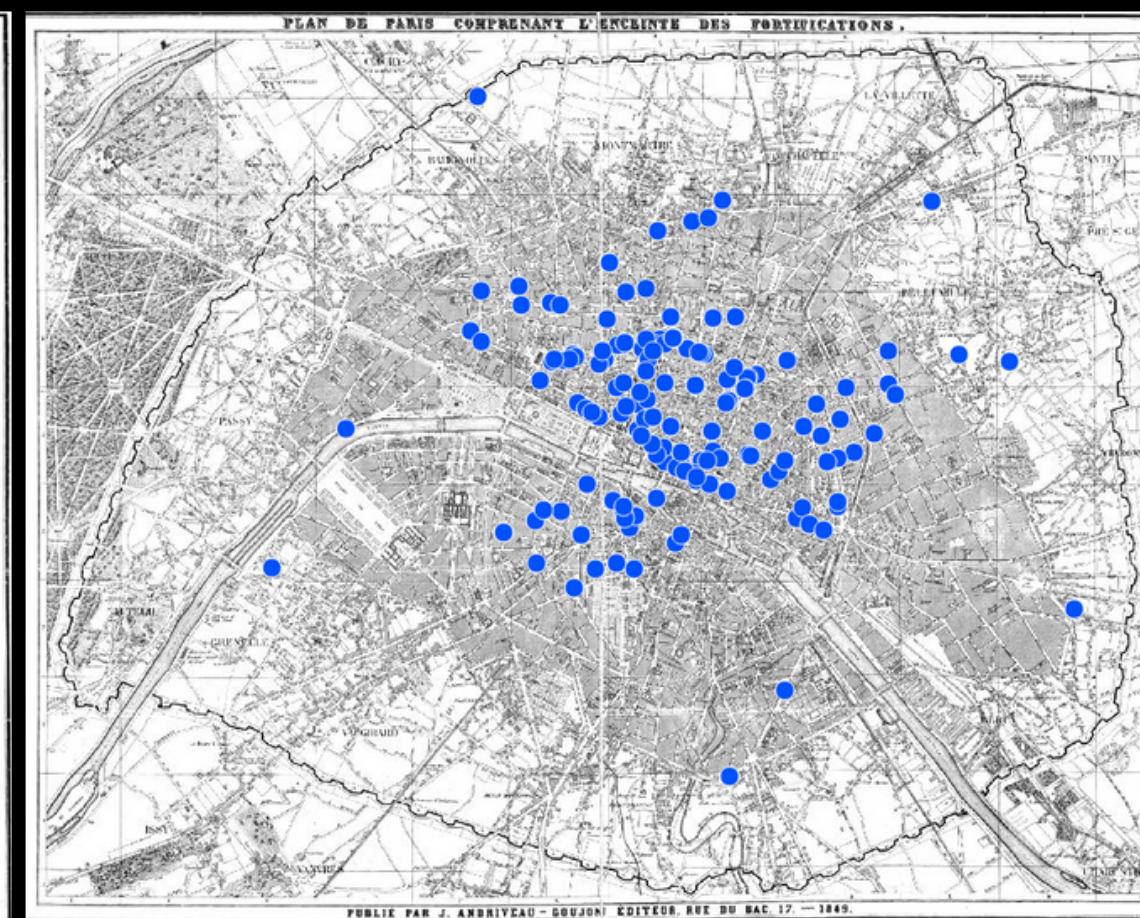
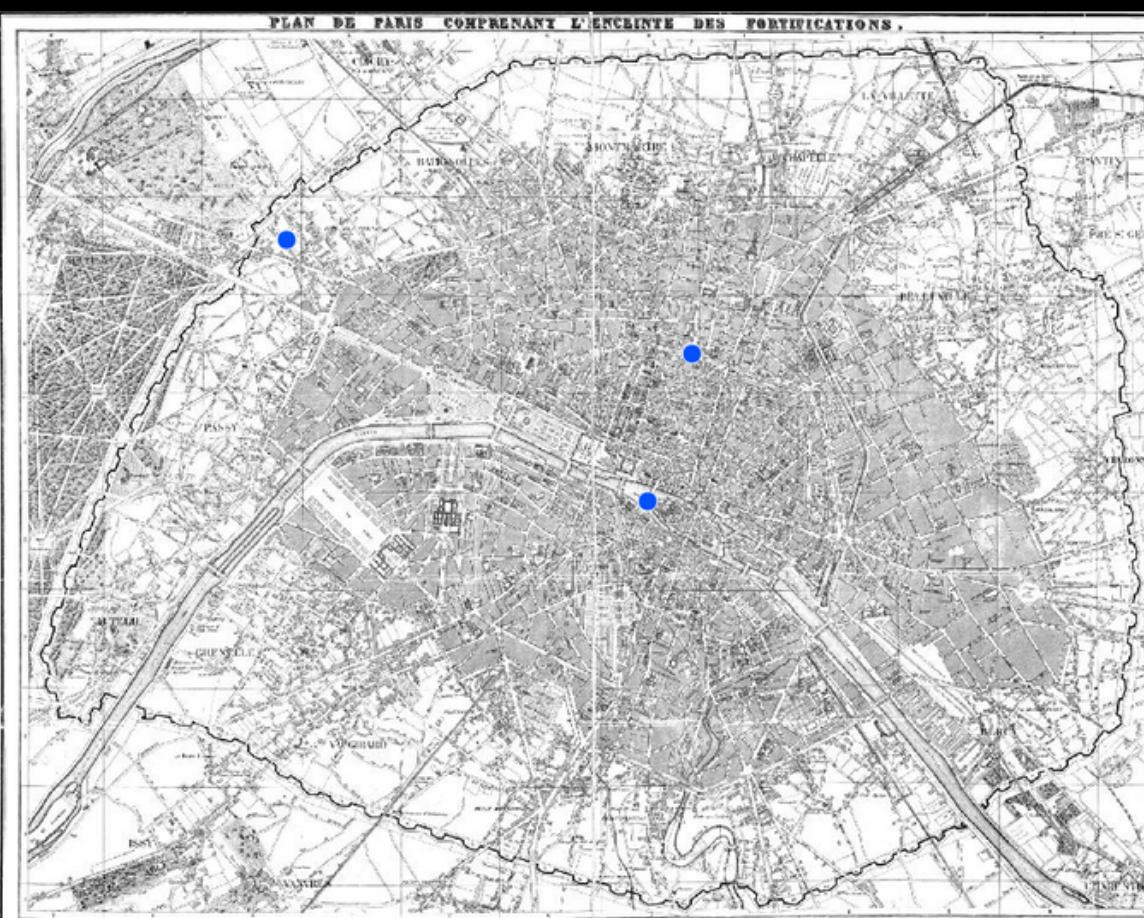


1875



daguer

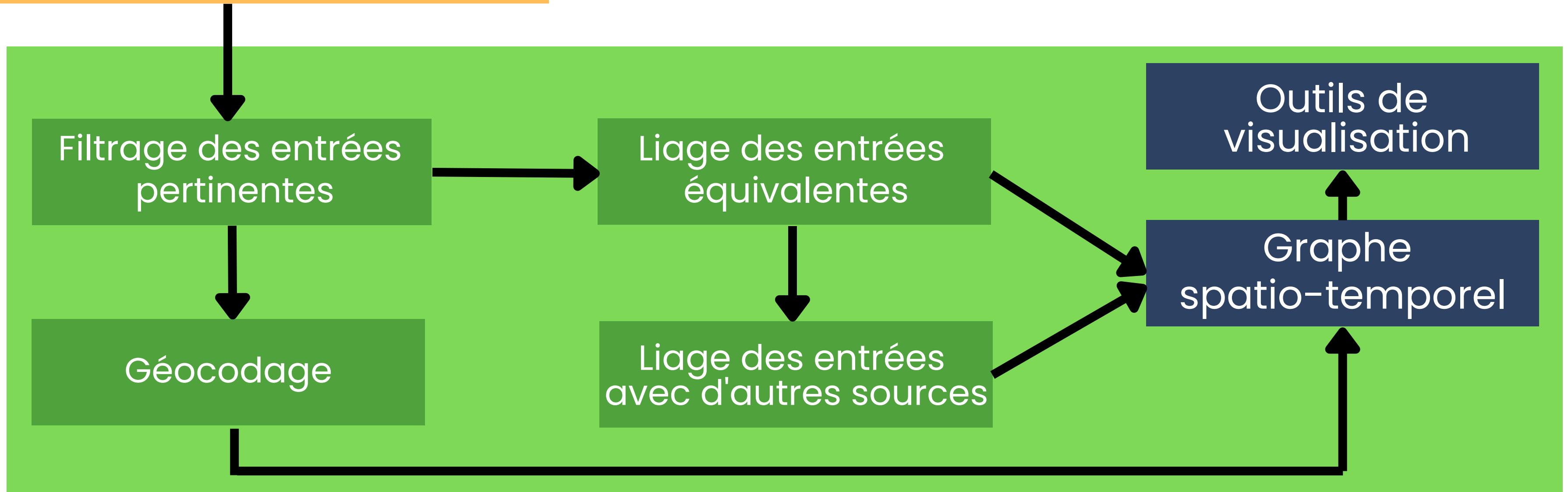
photo



## Conclusion

- Démarche fonctionnelle d'exploitation des données extraites des annuaires
- Reproductible pour d'autres cas d'applications

Entrées extraites et structurées



**Merci pour votre attention**



## Extraction de données

Nathalie Abadie, Edwin Carlinet, Joseph Chazalon et Bertrand Duménieu (mai 2022).  
« **A Benchmark of Named Entity Recognition Approaches in Historical Documents Application to 19th Century French Directories** ».  
Document Analysis Systems : 15th IAPR International Workshop, DAS 2022, La Rochelle, France.

## Liage de données et graphes de connaissances

### **Sakey : outil de détection de clés**



### **Silk Linked Data Integration Framework : outil de liage de données**

<https://github.com/silk-framework/silk>

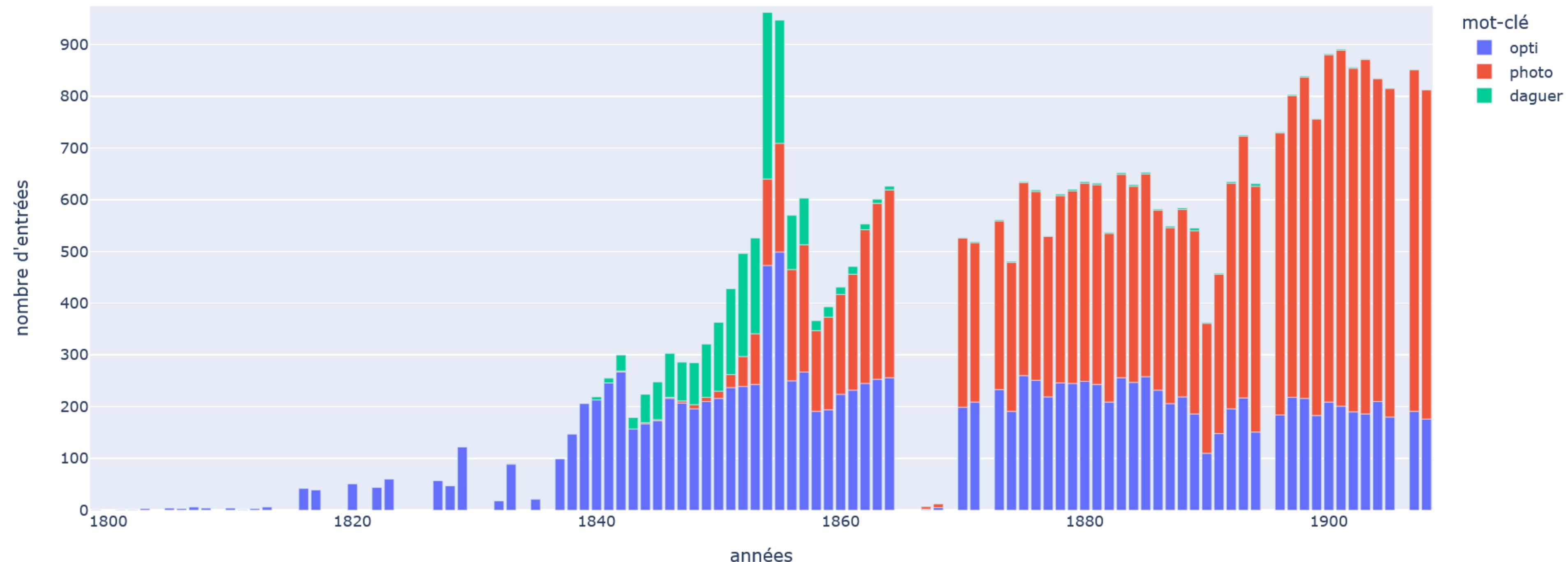
## Ouvrage de référence sur les photographes

Marc Durand (2015). « **De l'image fixe à l'image animée: 1820-1910. Tome 2: actes des notaires de Paris pour servir à l'histoire des photographes et de la photographie** ». Archives nationales. Pierrefitte-sur-Seine.

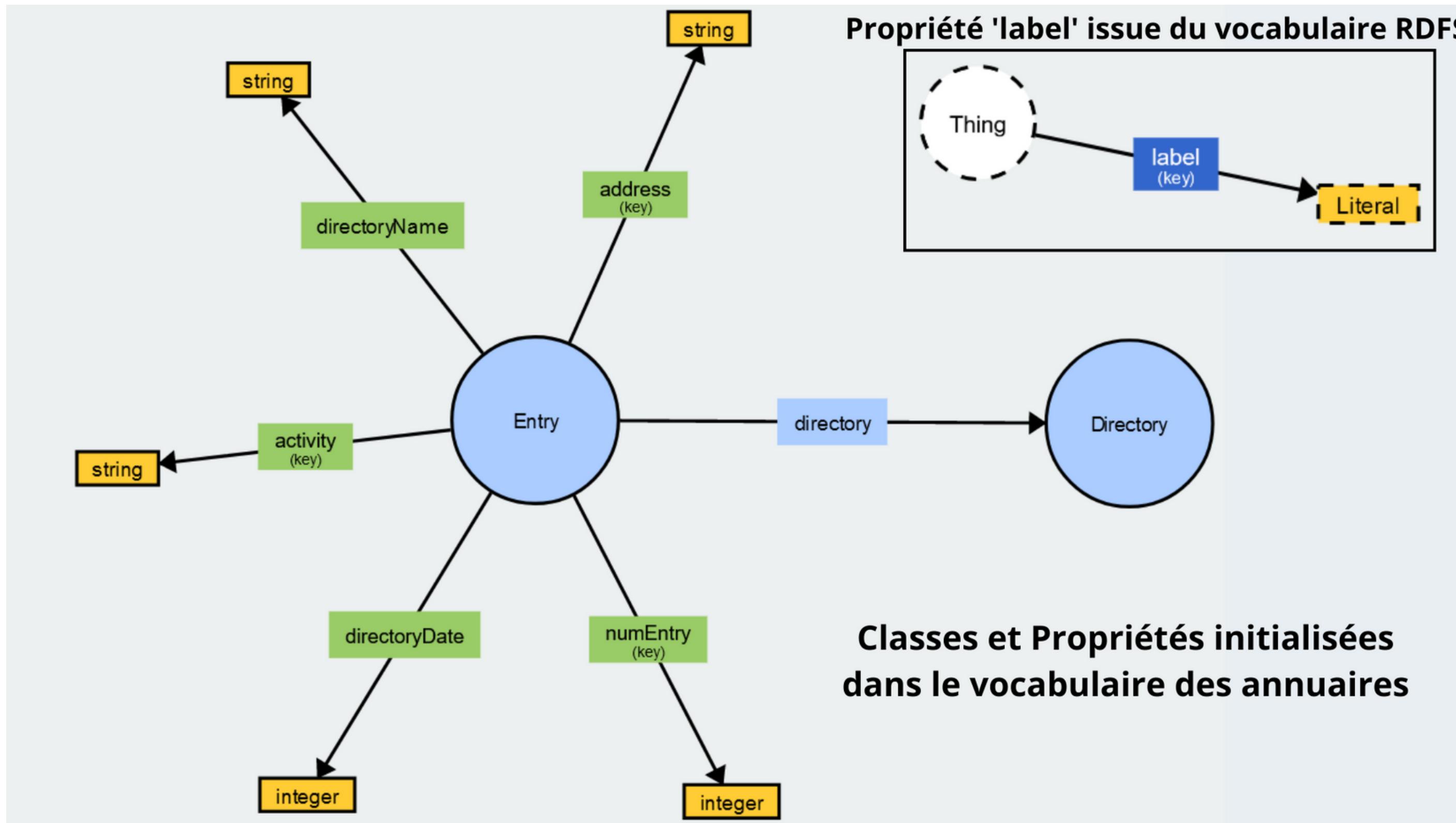
# Annexe : Détail des entrées extraites

## 3.1. Filtrage des entrées

Evolution du nombre d'entrées contenant les mots-clés 'photo','daguer' et 'opti' par année (table : par\_activites)



# Annexe : Ontologie utilisée pour le raisonnement - Méthode logique



## Annexe : Requête vers DATA BNF

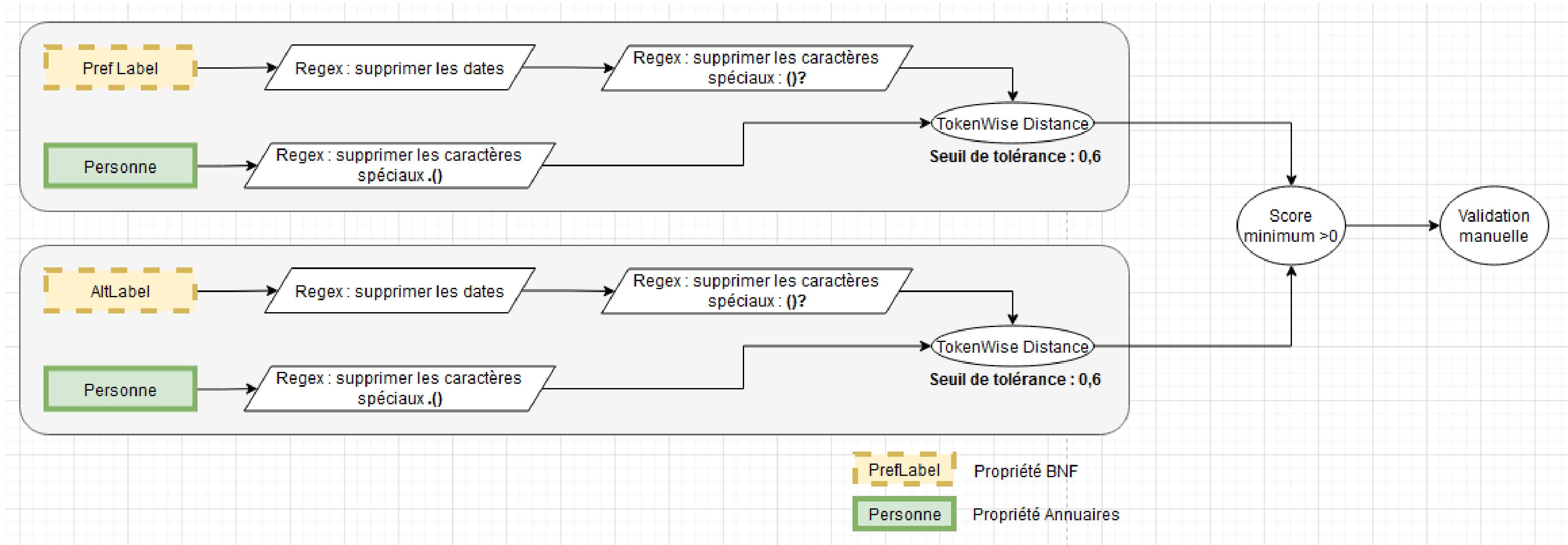
```
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX bnf-onto: <http://data.bnfr/ontology/bnf-onto/>
CONSTRUCT {
    ?s a foaf:Person;
    bnf-onto:firstYear ?fy ;
    bnf-onto:lastYear ?ly ;
    skos:prefLabel ?pf;
    skos:altLabel ?al.
}
WHERE {
    ?c a skos:Concept; skos:prefLabel ?pf; skos:altLabel ?al; foaf:focus ?s.
    ?s a foaf:Person ;
    bnf-onto:firstYear ?fy ;
    bnf-onto:lastYear ?ly ;
    rdagroup2elements:countryAssociatedWithThePerson <http://id.loc.gov/vocabulary/countries/fr> ;
    rdagroup2elements:fieldOfActivityOfThePerson <http://dewey.info/class/770/>, "Photographie" .
    Filter (?fy > 1760 && ?fy < 1885)
}
```



<https://data.bnfr/sparql>

76 fiches

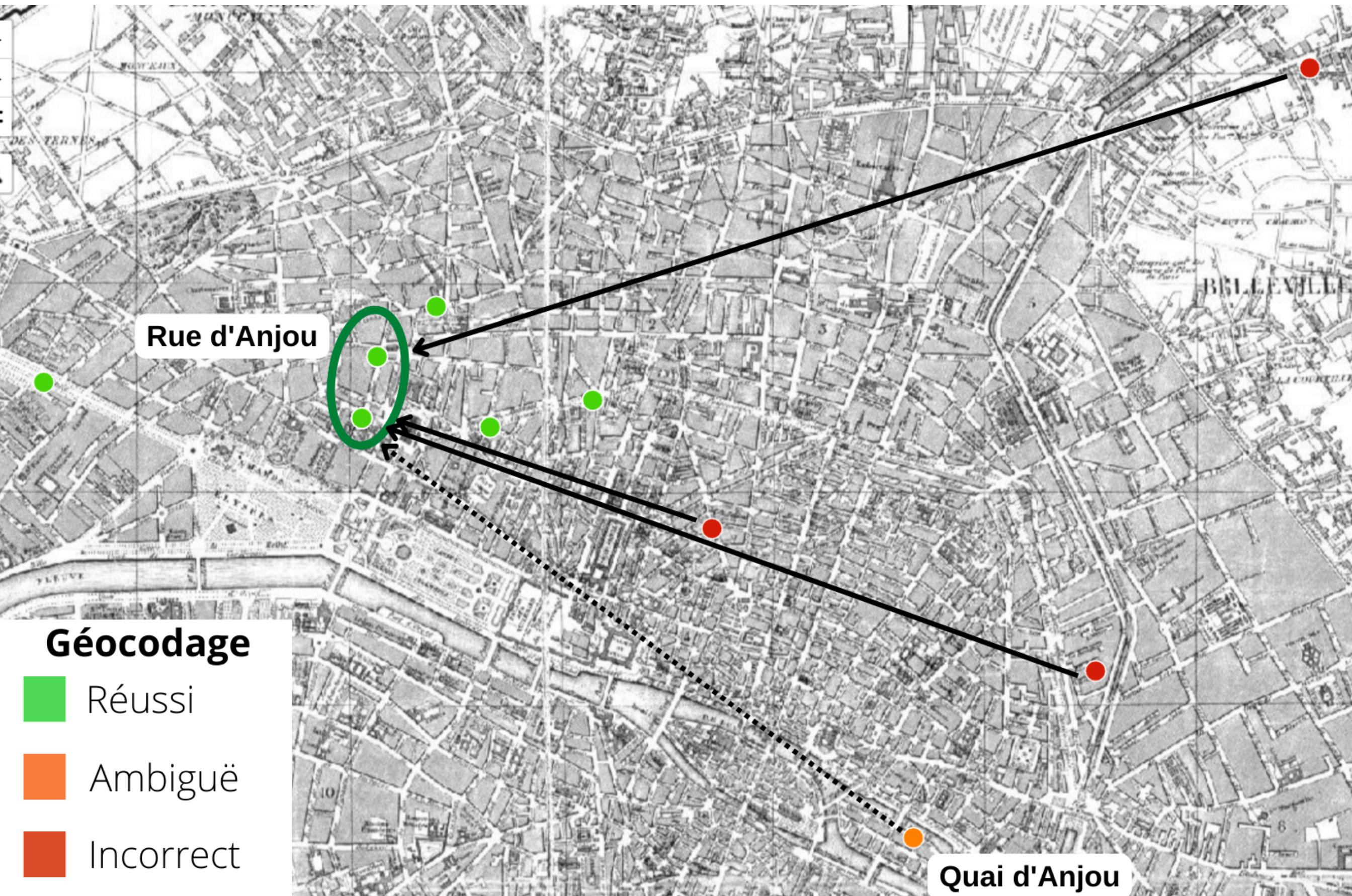
## Annexe : Appariement Annuaires - Data BNF



161 liens validés vers 29 fiches

## Annexe : Problématiques liées au géocodage

### Visualisation cartographique



Ateliers de la famille  
Tournachon - Nadar  
entre 1850 et 1908



Demo V1 en ligne  
[https://solenn-tl.github.io/  
stage\\_demo\\_photographies/](https://solenn-tl.github.io/stage_demo_photographies/)