

# Création de graphes de connaissances géohistoriques à partir des annuaires du commerce parisien du 19ème siècle

Solenn Tual <sup>(1)</sup>, Nathalie Abadie <sup>(1)</sup>, Bertrand Duménieu <sup>(2)</sup>, Joseph Chazalon <sup>(3)</sup>, Edwin Carlinet <sup>(3)</sup>

*L'Intelligence artificielle dans le traitement d'images numériques en SHS,  
MSH Val de Loire (Université de Tours) et Criham (Université de Poitiers), 27 novembre 2023*

*Licence : CC BY 4.0*

# Annuaires du commerce de Paris

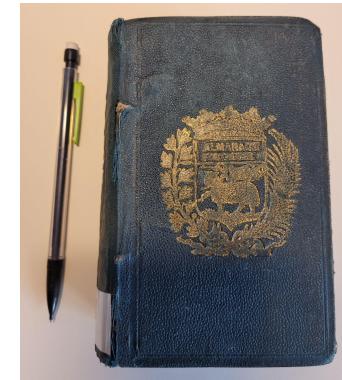
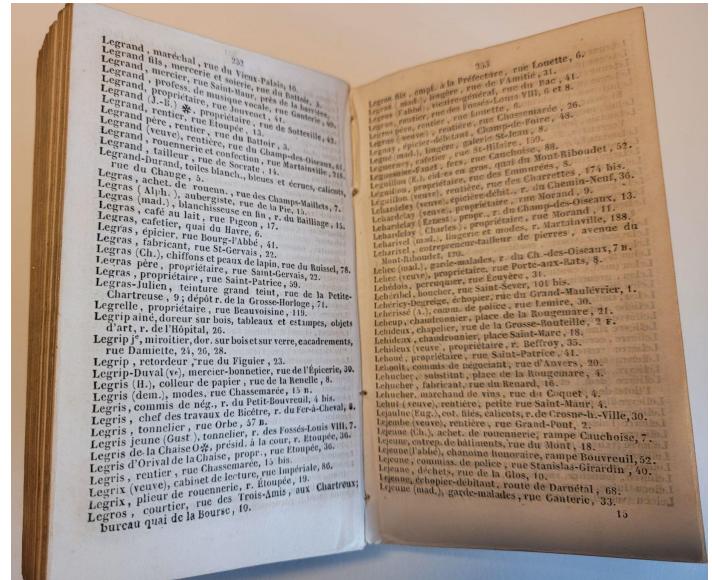
- Édités à partir de la fin du XVIIIème siècle jusqu'au milieu du XXème siècle par différents éditeurs.
- Organisés en différentes listes (alphabétique, par profession, par adresse).
- Chaque entrée contient :
  - Nom/Raison sociale
  - Activité/Profession/Statut
  - Adresse(s)
  - Distinctions militaires/professionnelles

Cherot, joaillier. S. Martin, 51.  
 Chéroux, orf. Ste 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 de 1839.

Source : Gallica BNF



# Annuaires du commerce de Paris

- Redondance importante des entrées d'une année ou d'une édition à une autre
- Snapshots temporels

même personne

Bibliothèque Ste-Geneviève, 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, fab. de jouets, Montmorency, 33.
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.

succession

Bibliothèque Ste-Geneviève, 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, fab. de jouets, Montmorency, 33.
Bical et Dorre, fab. de socques, Vert-Bois, 14.
Bican (Vve) et fils, fondeur en cuivre , place de la Corderie-du-Temple, 26.
Bicel, épicier, marché d'Aguesseau, 15.
Bichard (Mme), Nve-de-Luxembourg, 17.
Bichard, tabacs et eau-de-vie , Faub.-St-Martin, 45.

déménagement

Bibliothèque Ste-Genevière, 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 (Mme), Nve-de-Luxembourg, 17.
Bichard, tabacs et eau-de-vie , Faub.-St-Martin, 45.

Bibliothèque Ste-Genevière, rue des Sept-Voies et place du Panthéon.
Bibliothèque de la Ville , quai d'Austerlitz , 33 (provisoirement).
Bibolet, relieur, passage Sainte-Marie-Saint-Germain, 10.
Bibonne, architecte, Magasins, 12.
Bibron, aide-naturaliste au Jardin-des-Plantes, Cuvier, 29.
Bibus, tailleur, Richelieu, 31.
Bical, fab. de jouets, Montmorency, 33.
Bical (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

Didot 1844a - pages 125-126

Source images : Gallica BNF

# Des sources géographiques anciennes à l'analyse géohistorique

Reconnaissance

Classification

Géo-référencement

Structuration

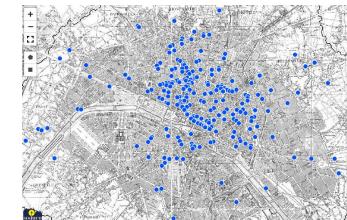
Scan → Texte

Ravrio et comp., fabr. de bronzes et cu-  
riosités, r. Richelieu, 93; la fabrique  
rue Montmartre, 161.

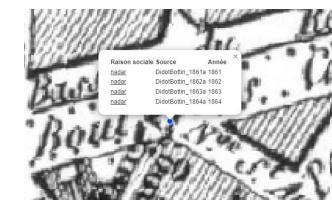
Texte →  
Entités nommées spatiales

Ravrio et comp. PER , fabr. de bronzes et cu-riosités ACT ,  
r. Richelieu LOC , 93 CARDINAL ; la fabrique FT  
rue Montmartre LOC , 161 CARDINAL .

Entités nommées spatiales →  
Données géographiques



Données géographiques →  
Données géohistoriques



Segmentation des pages  
+  
Optical Character  
Recognition (OCR)

Reconnaissance des  
Entités Nommées (NER)

Géocodage

Filtrage et liage des  
entrées

# Extraction et structuration des entrées

## Segmentation des entrées

Pré-traitement image  
(correction géométrique,  
réduction de bruit...)



Segmentation du canvas (x-y cut)  
Classification des blocs  
Détection des lignes (watershed), fusion des lignes appartenant à  
une même entrée)



Backes (F.), succ. de F. Delacour <sup>1811</sup> et Backes, Elzévir, 7.

Barré (Gustave), armes blanches de l'art, sabres, épées de combat, couteaux de chasse; réparat. d'armes antiques; comm. export., Lions-St-Paul, 11.

Delacour (F.) <sup>1811</sup> et Backes (F. Backes, success.), fabr. d'armes blanches, 1867, Elzévir, 7.

Fauré Le Page <sup>anc. malson Le Page</sup>, rue Bochelle, 8.

Gambette (G.), armures, armes blanches, copies d'épées anciennes, Quatre-Septembre, 2.

Grain (L.), armes blanches, spds. de fourreaux et garnitures de sabres, r. Beaubourg, 6.

Grégoire (A.), Écouffes, 20.

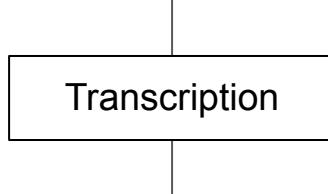
Lacloix (L.), armes françaises et étrangères, couteaux de chasse, rue de Thorigny, 6.

Maria (J.), rue du Quatre-Sep-tembre, 14.

Petitfils (W.), Ecouffes, 22.

# Reconnaissance optique des caractères (OCR)

Lemonnyer, plomberie, r. de Bondy, 86, et  
r. Bouchardon, 1.



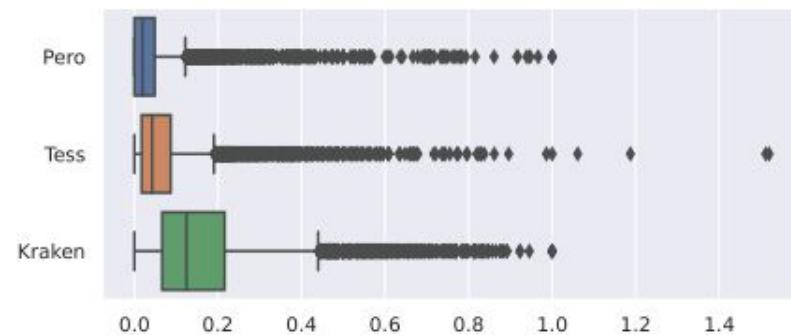
Lemonnyer, plomberie, r. de Bondy, 86, et  
r. Bouchardon, 1.

Dataset DOI [10.5281/zenodo.6394464](https://doi.org/10.5281/zenodo.6394464)

Comparaison de 3 modèles d'OCR

	PERO OCR	Tesseract	Kraken
CER	3.78%	6.56%	15.72%

CER : Character Error Rate



N. Abadie, E. Carlinet, J. Chazalon et B. Duménieu, "A Benchmark of Named Entity Recognition Approaches in Historical Documents Application to 19th Century French Directories", in Proc. DAS 2022,  
<https://github.com/soduco/paper-ner-bench-das22>

# Reconnaissance des entités nommées (NER)

### Datasets

PER, ACT, LOC, CARD



**Dataset Référence** : entrées transcris et annotées manuellement

→ Dufant (Victor), libraire, r. du Gros-Chenet, 2. 392  
TDufant (Victor), libraire, r. du Gros-Chenet, 2.JO



**Dataset Bruité** : mêmes entrées transcris avec PERO-OCR et annotations alignées automatiquement

→ Dutils, limonadier, r. des Colonnes, 6. 395  
lukils. limonauier, 1. des Colonnes, 6. J9

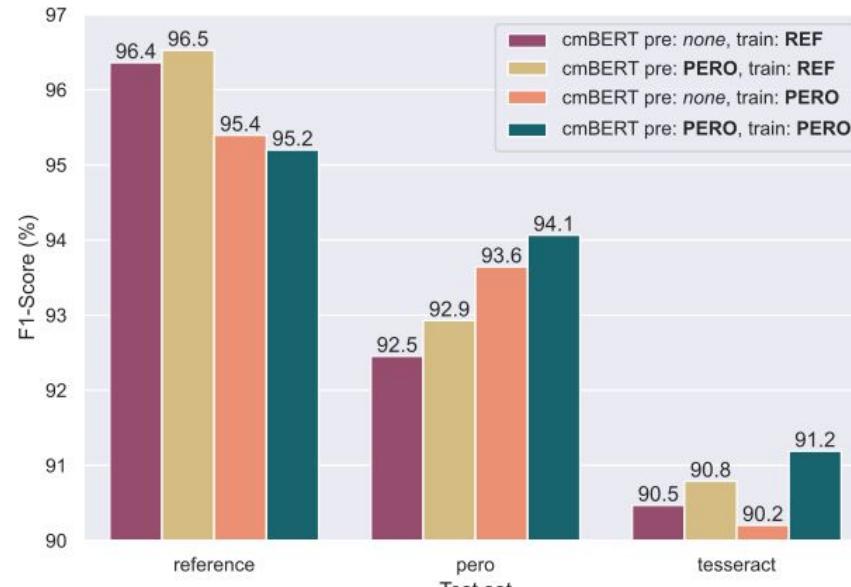
### Modèles (Transformers)

- CamemBERT
- CamemBERT pré-entraîné avec des entrées d'annuaires OCrisées

**8341** entrées annotées :

- Train : 6004 entrées
- Dev : 668 entrées
- Test : 1669 entrées

# Reconnaissance des entités nommées (NER)



Abadie et al. (2022)

Mettereauprop., quai d'Anjou, 7.\nMettemberg, élig., méd., St-Thomas-d'Enf., 5.\nMetz (de), rentier, St-Guillaume, 30.\nMetzinger, avocat, Rameau, 6.\nMetzmacher, peint. sur émaux, St-Martin, 124.\nMeurgey, épicier-herboriste, Dragon, 33.\nMeurice, Chaussée-d'Antin, 3.\nMeurice (Eug.), tapissier, Vivienne, 12.\nMeurillon, marbrier-sculpteur, butte Mont-\nParnasse, 15.\n



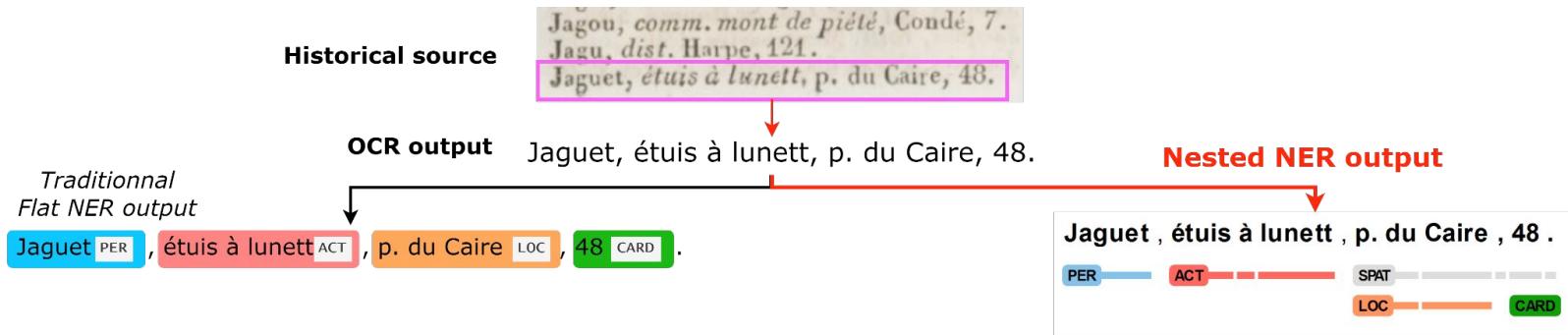
Mettereauprop., quai d'Anjou LOC, 7 CARDINAL.  
Mettemberg élig. TITRE, méd. ACT, St-Thomas-d'Enf. LOC, 5 CARDINAL.  
Metz (de) PER, rentier ACT, St-Guillaume LOC, 30 CARDINAL.  
Metzinger PER, avocat ACT, Rameau LOC, 6 CARDINAL.  
Metzmacher PER, peint. sur émaux ACT, St-Martin LOC, 124 CARDINAL.  
Meurgey PER, épicier-herboriste ACT, Dragon LOC, 33 CARDINAL.  
Meurice PER, Chaussée-d'Antin LOC, 3 CARDINAL.  
Meurice (Eug.) PER, tapissier ACT, Vivienne LOC, 12 CARDINAL.  
Meurillon PER, marbrier-sculpteur ACT, butte Mont-Parnasse LOC, 15 CARDINAL.



Base de données contenant environ 22 millions d'entrées structurées

=> Environ 10 millions avec la première extraction

# Test non implanté : Reconnaissance des entités nommées imbriquées (NER)



- Nouveaux datasets (ocr brut et ocr corrigé manuellement, mêmes entrées que *Abadie et al.*) :
  - 2 niveaux d'entités
- Benchmark de 3 méthodes de NER
  - Respect de la hiérarchie des classes
  - Performances inchangées par rapport au NER “classique”

S. Tual, N. Abadie, E. Carlinet, J. Chazalon, et B. Duménieu, "A Benchmark of Nested NER Approaches in Historical Structured Documents", in Proc. ICDAR 2023, <https://github.com/soduco/paper-nestedner-icdar23-code/>

## Géocodage

### Restructuration des adresses

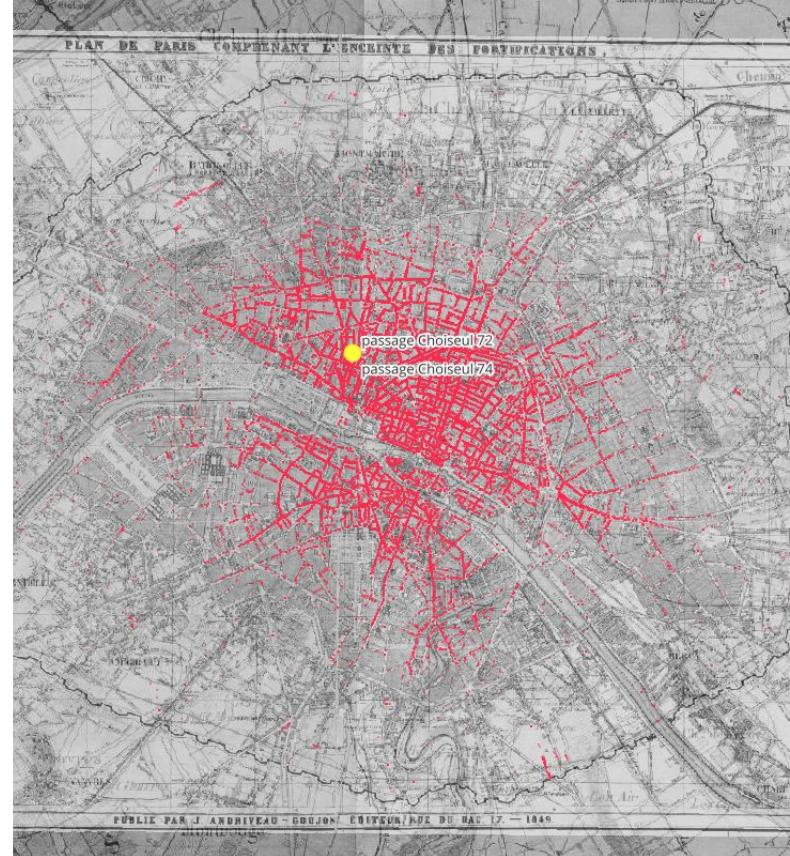
Aaron, bronzes et pendules,  
passage Choiseul, 72 et 74.



Aaron, bronzes et pendules,  
passage Choiseul, 72 et 74.

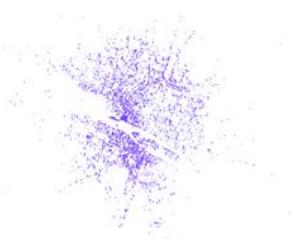


Adresse 1 : passage Choiseul 72  
Adresse 2 : passage Choiseul 74



## Géocodage

Sources géohistoriques des données d'adresses et de rues utilisées pour localiser les annuaires :



Rues de l'Atlas de Verniquet et  
adresses "Paris artistique,  
1800-1820"



Rues et adresses de l'atlas  
Jacoubet, 1827-1836



Rues et adresses de l'atlas  
municipal, 1888

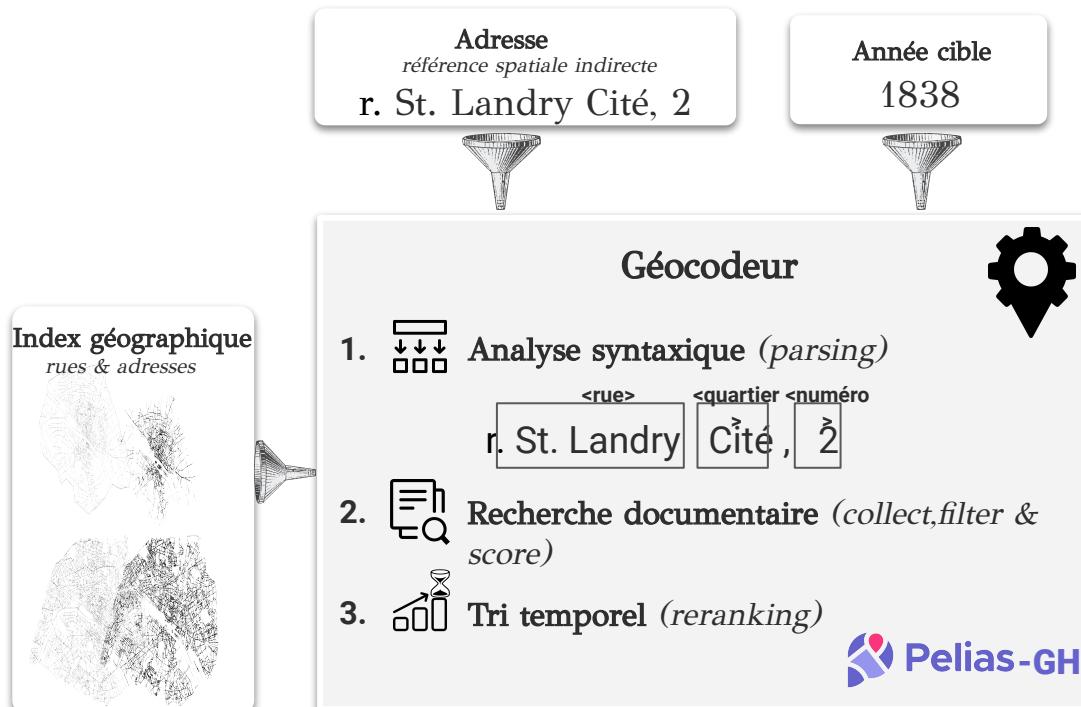


Rues et adresses OpenStreetMap, 2020

Duménieu (2023)

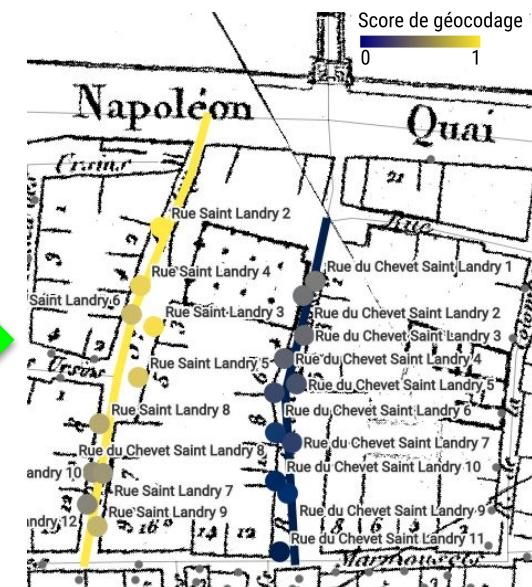
## Géocodage

Géocodeur sensible à la dimension temporelle



Outil open-source : [https://github.com/GeoHistoricalData/historical\\_geocoding](https://github.com/GeoHistoricalData/historical_geocoding)

96% des entrées sont  
géocodées.



## Questions de compétences

- ★ CQ1. Quelle est l'adresse du commerce X en 1862 ?
- ★ CQ2. Combien y a-t-il de commerces de ce type localisé *rue de Rivoli* en 1856 ?
- ★ CQ3. Quels sont les commerces situés dans une zone définie par un polygone ou un rectangle englobant en 1875 ?
- ★ CQ4. Quels commerces ont potentiellement déménagé au cours de leur existence ?
- ★ CQ5. Quels commerces ont été repris par un autre commerçant exerçant la même activité ?

## Sélection des entrées 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 dans la base de donnée des extractions

Vivienne (36, rue)
GUILLOT-SAGUZ (Amélie) [1849-1856]
PLUMIER (famille) [1856-1865] avec ROUGE (Antoine) dit TONY LE ROUGE
CHAPDELET (Charles Louis) dit KELARY [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]

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.



```
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

## Sélection des entrées pertinentes

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

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
Lithographie	litho	Prisme	prisme
Artiste	artiste	Appareil/instrument de mesure	mesure
Chambre noire	chamb AND noir	Camera obscura	camera
Cinéma, cinématographe	cinem		

- Vocabulaire trop large, introduit trop d'entrées qui ne sont pas relatives à la profession étudiée
- Juin 2022 : ~250 000 entrées

## Sélection des entrées 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.

<b>PHOTO</b>	<b>DAGUER</b>	<b>OPTI</b>
Photographe	Daguerréotype	Optique
Photographie		Opticien



Juin 2022 : 34 062 entrées



UPDATE Octobre 2023 : 45 807 entrées (listes alphabétiques uniquement)

Création de ressources RDF à partir de la base de donnée relationnelle  
contenant ces entrées (avec *ontop*)

## Sélection des entrées 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.

<b>PHOTO</b>	<b>DAGUER</b>	<b>OPTI</b>
Photographe	Daguerréotype	Optique
Photographie		Opticien



Juin 2022 : 34 062 entrées



UPDATE Octobre 2023 : 45 807 entrées (listes alphabétiques uniquement)

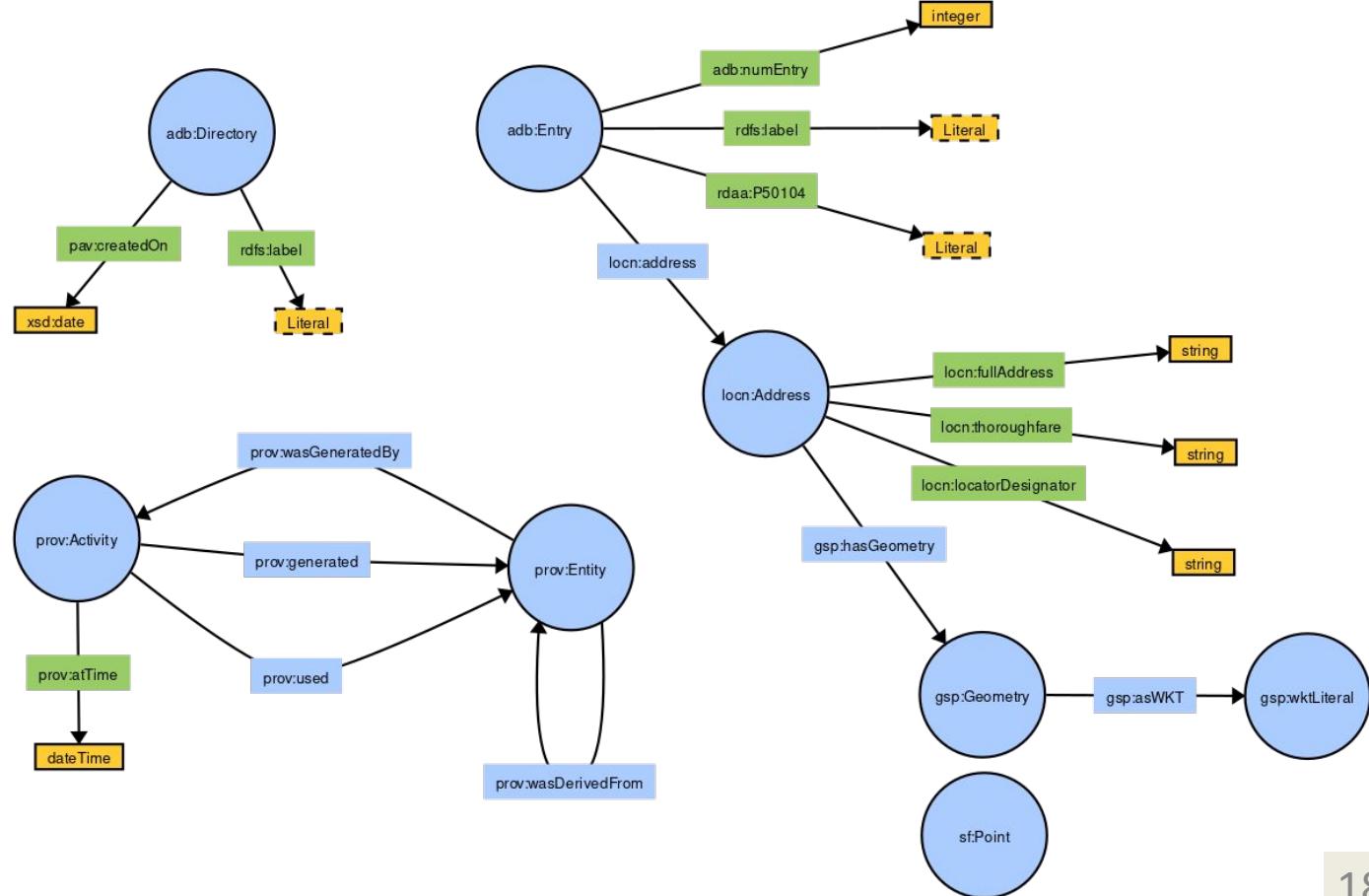
Création de ressources RDF à partir de la base de donnée relationnelle  
contenant ces entrées (avec *ontop*)

# Construction du graphe de connaissances géohistorique

## • Ontologie

### Vocabulaires utilisés :

- **RDFS** : label des entrées
- **LOCN** : description des adresses
- **Geosparql** : description des objets géographiques
- **RDA Registry** : description des activités
- **PROV** : description des données



- Méthodes de liage

## Méthode fondée sur les connaissances

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

→ Clé = combinaison de propriétés, découvertes avec Sakey

- Numéro de l'entrée (si plusieurs activités ou adresses dans l'entrée)
- Nom et Activité
- Nom et Adresse
- Adresse et Activité

→ Création des liens par inférence

- Méthodes de liage

## Méthode fondée sur les données

- Appariement par **comparaison numérique** des propriétés
- Distances d'édition : Levenshtein, TokenWise

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

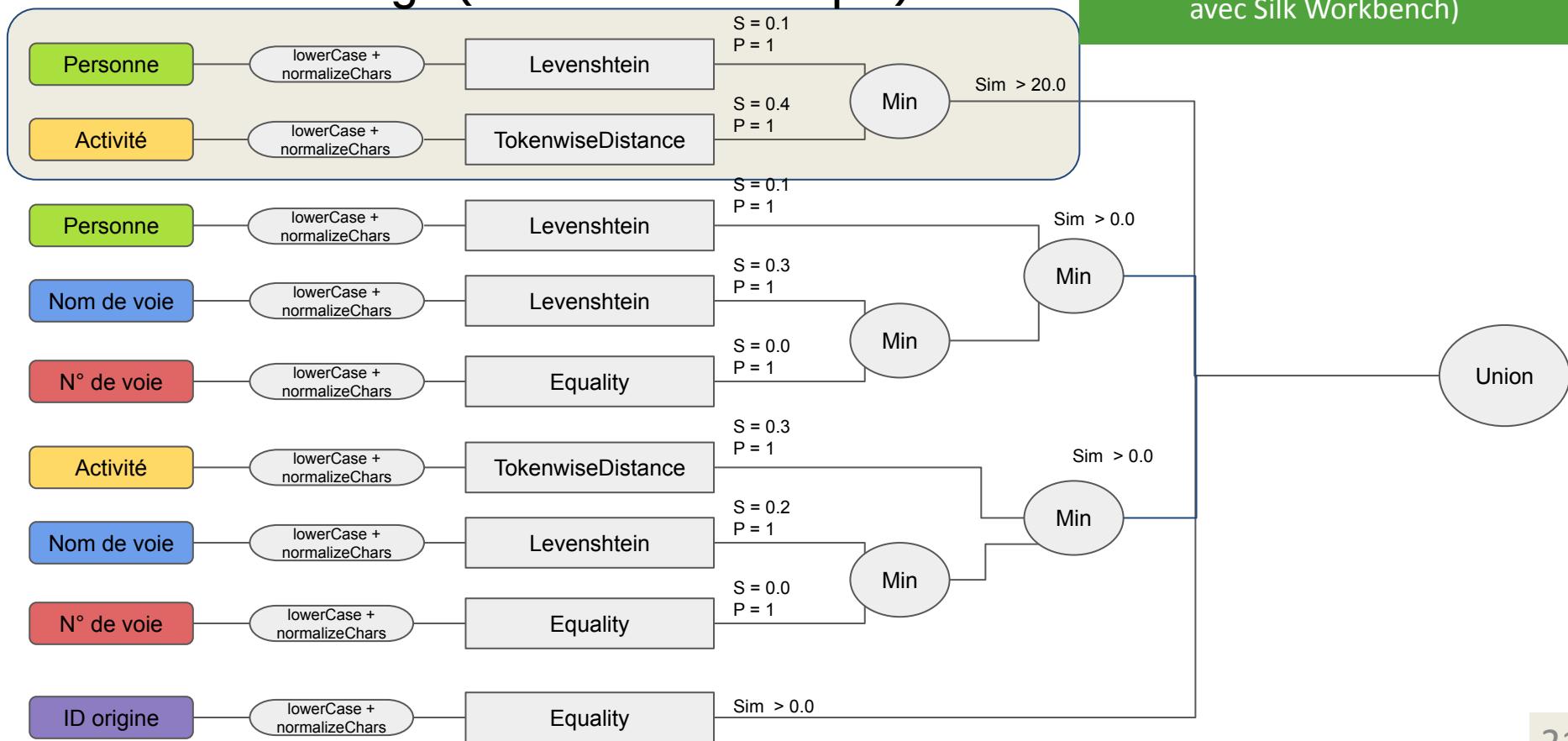
ex :    **CHAT** <=> **PLAT**        2 remplacements

Distance TokenWise :

- Calcul de la distance de Levenshtein entre tous les mots du texte
- Score final normalisé (entre 0 et 1)

# Construction du graphe de connaissances géohistorique

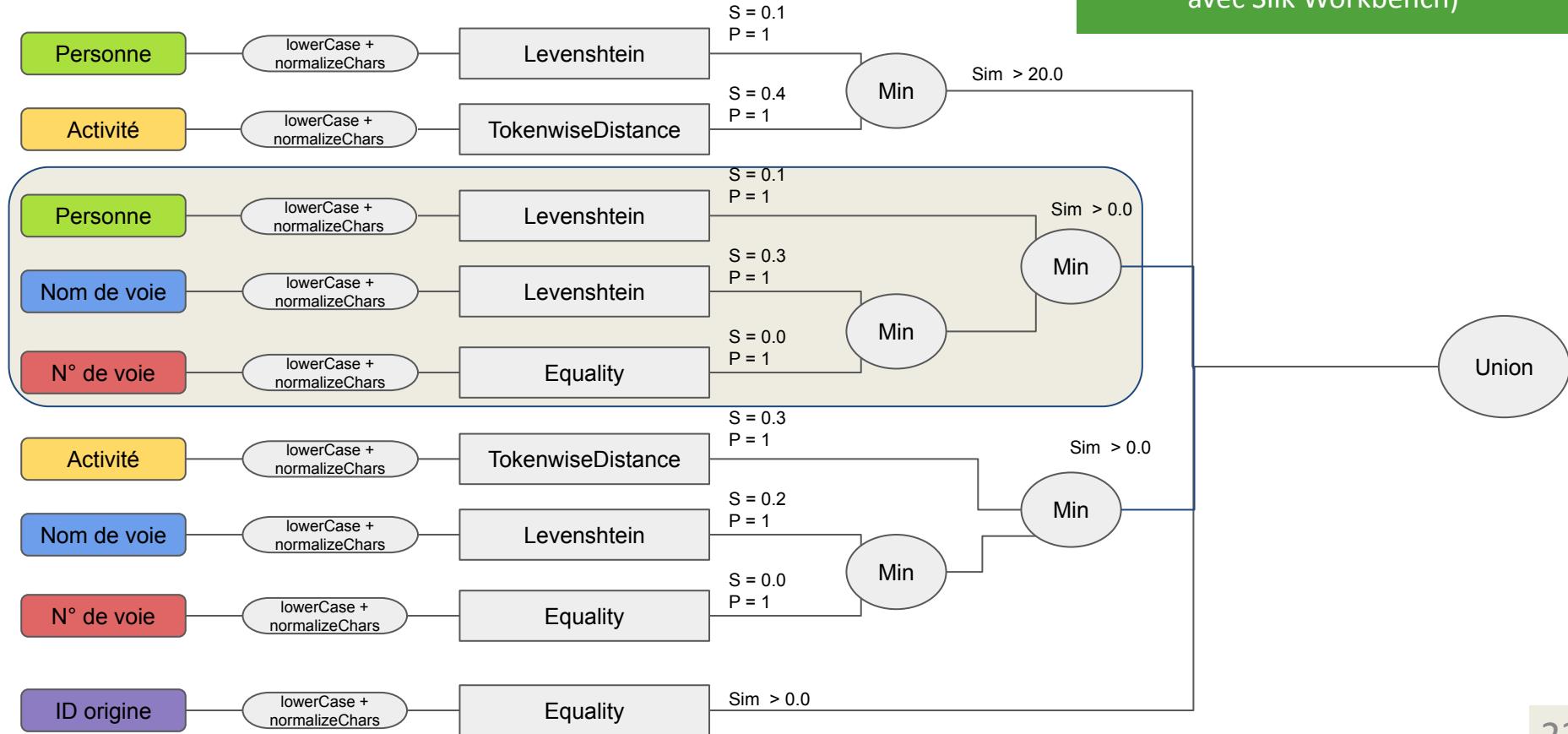
## • Critères de liage (méthode numérique)



# Construction du graphe de connaissances géohistorique

## • Critères de liage (méthode numérique)

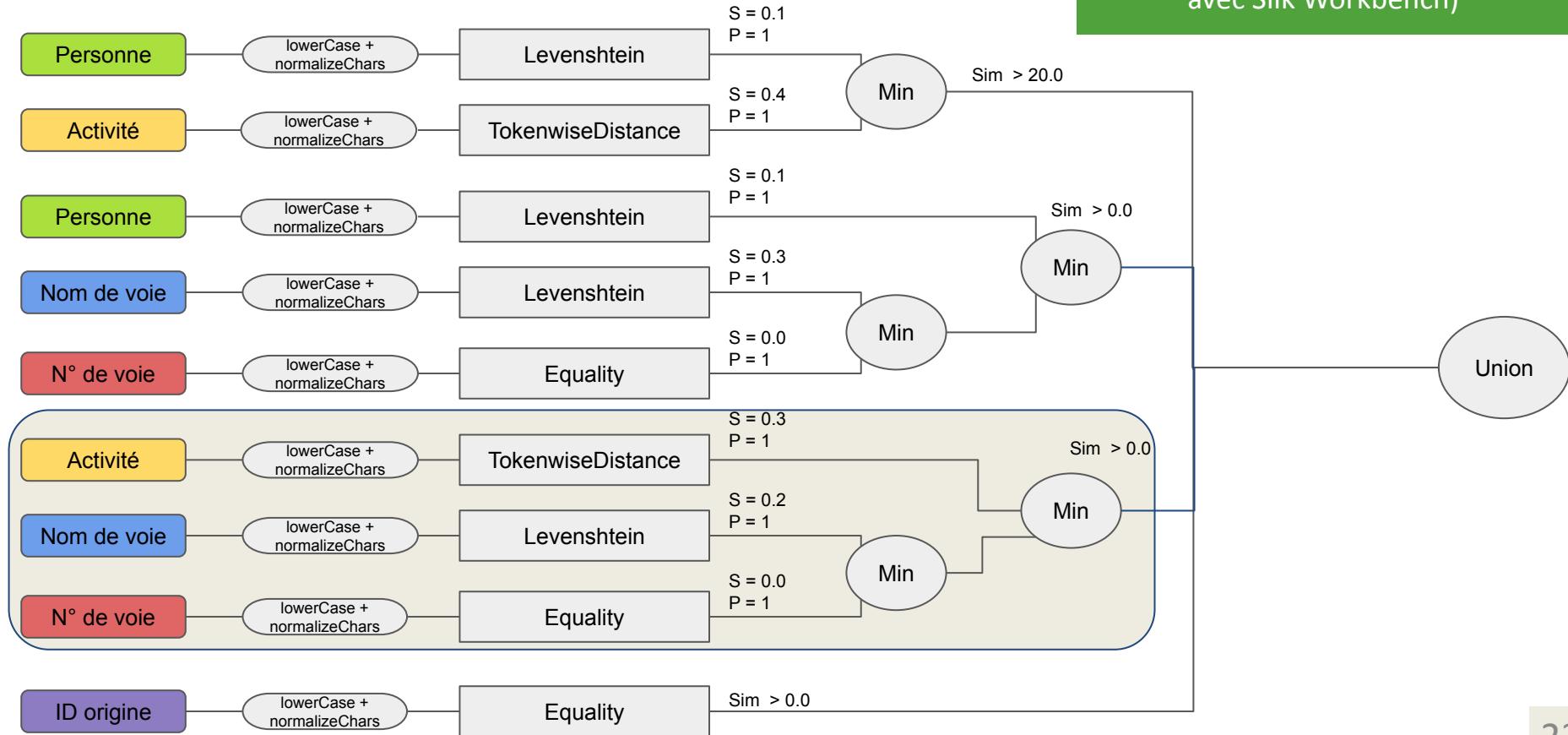
Liage réalisé avec Silk  
Single-machine (tests des critères avec Silk Workbench)



# Construction du graphe de connaissances géohistorique

## • Critères de liage (méthode numérique)

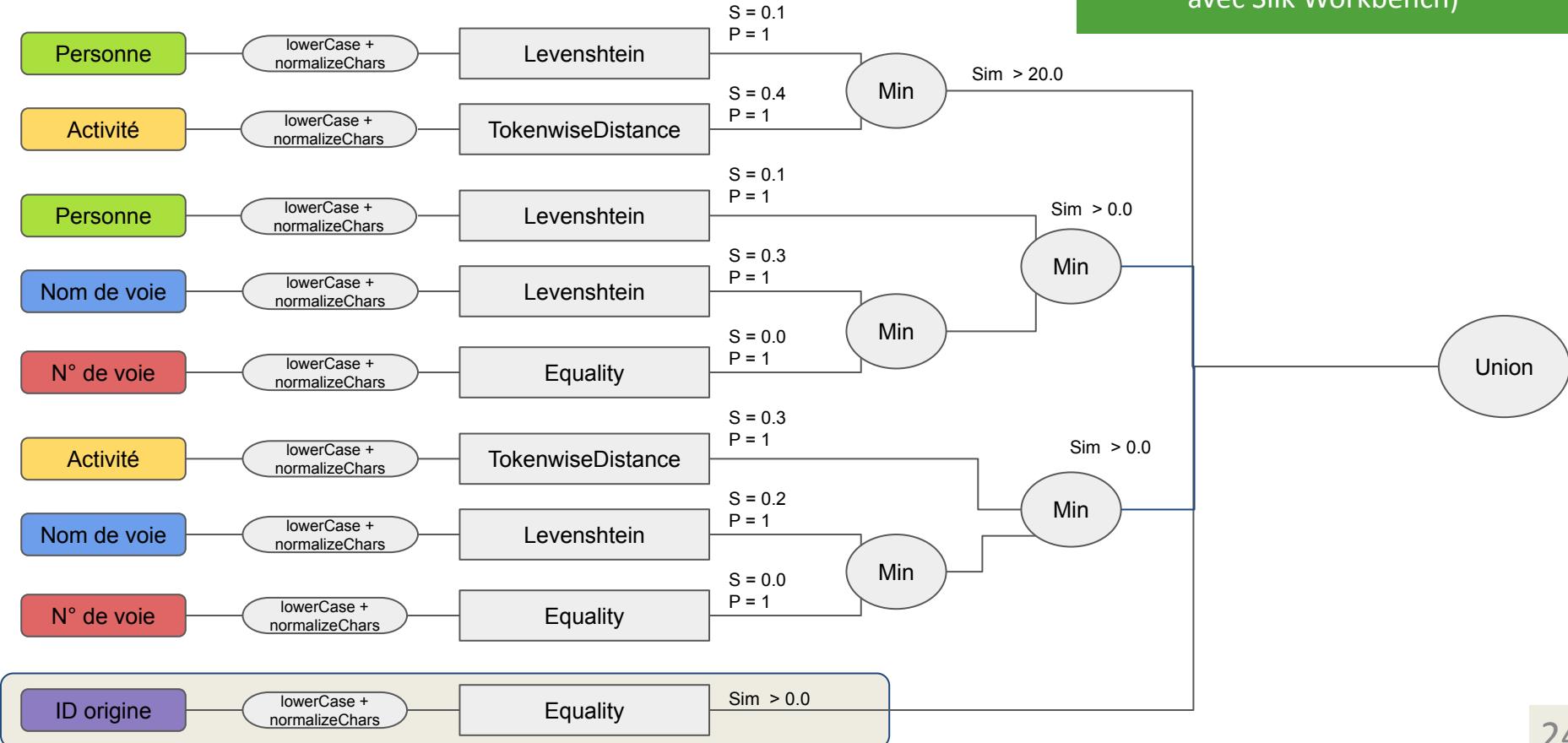
Liage réalisé avec Silk  
Single-machine (tests des critères avec Silk Workbench)



# Construction du graphe de connaissances géohistorique

## • Critères de liage (méthode numérique)

Liage réalisé avec Silk  
Single-machine (tests des critères avec Silk Workbench)



- Méthodes de liage

## Méthode logique

Paramétrage simple = clés

Comparaison très stricte :

- peu adaptée pour traiter des textes bruités
- liens sûrs
- temps et ressources de calcul importantes

1ère extraction

34 062 entrées



250 622 liens

## Méthode numérique

Paramétrage complexe = identifier les seuils de tolérance pertinents

Comparaison numérique :

- plus tolérante au bruit OCR
- risque plus élevé de produire des liens erronés (importance de choisir des seuils plutôt strictes)

& Propagation des liens sameAs

34 062 entrées



357 130 liens

- Méthodes de liage

## Méthode numérique

Paramétrage complexe = identifier les seuils de tolérance pertinents

Comparaison numérique :

- plus tolérante au bruit OCR
- risque plus élevé de produire des liens erronés (importance de choisir des seuils plutôt strictes)



Graphe géohistorique (SPARQL Endpoint) :  
<https://dir.geohistoricaldata.org/sparql>

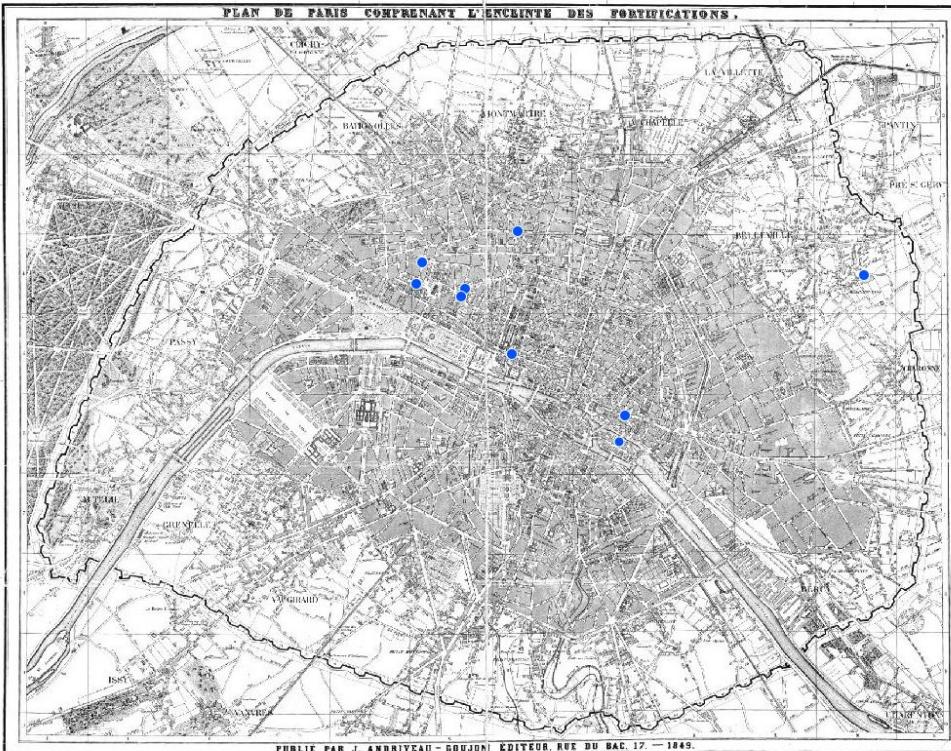
2ème  
extraction



# Visualisation et évaluation des questions de compétences

## Visualisation spatiale...

[bit.ly/graphes\\_geohistoriques\\_soduco](http://bit.ly/graphes_geohistoriques_soduco)



Plan de Paris contenant l'enceinte des fortifications - Andriveau-Goujon - 1849 © GeoHistoricalData

Visualisation des graphes géohistoriques construits à partir des entrées d'annuaires du commerce de Paris (XIXème siècle)

[SPARQL Endpoint](#) | [Dépôt Git-Hub](#) | [Aide](#)

### Dataset

Photographes et professions associées

Statistiques du dataset

### Filtres

#### Propriétés

##### Raison sociale

nadar

##### Description

Ex : photo

##### Adresse

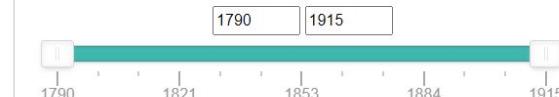
Ex : rivoli

Les champs du formulaire acceptent les expressions régulières. Deux regex utiles :

- La propriété contient mot1 ET mot2: (?=.+"mot1)(?=.+"mot2)
- La propriété contient mot1 OU mot2: mot1|mot2

[Tester une regex complexe.](#)

#### Période



Le filtre temporel permet de faire varier l'affichage des points préalablement chargés sur la carte sans lancer une nouvelle recherche.

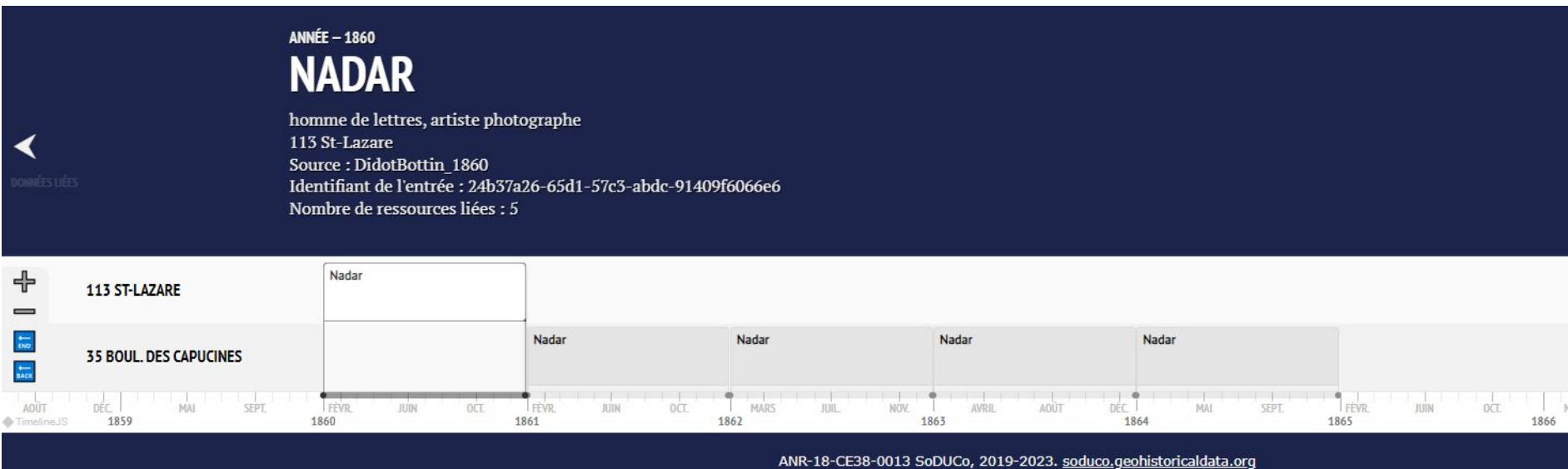
Données chargées pour la période 1790-1915.

# Visualisation et évaluation des questions de compétences

... et temporelle



[bit.ly/graphes\\_geohistoriques\\_soduco](http://bit.ly/graphes_geohistoriques_soduco)



- ★ CQ1. Quelle est l'adresse du commerce X en 1862 ?
- ★ CQ2. Combien y a-t-il de commerces de ce type localisé *rue de Rivoli* en 1856 ?
- ★ CQ3. Quels sont les commerces situés dans une zone définie par un polygone ou un rectangle englobant en 1875 ?
- ★ CQ4. Quels commerces ont potentiellement déménagé au cours de leur existence ?
- ★ CQ5. Quels commerces ont été repris par un autre commerçant exerçant la même activité ?

# Visualisation et évaluation des questions de compétences

## Questions de compétences

Quelle est l'adresse du commerce de Gallino en 1862 ?

11 rue suger

The screenshot shows a historical map of Paris from 1862. A blue dot marks the location of 'Gallino' at '11 Rue Suger'. A callout box provides the following details:

Gallino  
Adresses (annuaire) : 11 Suger  
Adresse (géocodage) : 11 Rue Suger, Paris (Source: atlas\_municipal\_1862)  
Activité : opticien, fab de baromètres et instruments de physique  
Année de publication : 1862  
Annuaire : Diderotin\_1862  
Identifiant de l'entrée : 8e5c8a38-c9ea-56c3-a930-17e51ae3f5e0

A large orange arrow points to the text '11 rue suger'.

**PREFIXES**

```
PREFIX locn: <http://www.w3.org/ns/locn#>
PREFIX ont: <http://rdf.geohistoricaldata.org/def/directory#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX pav: <http://purl.org/pav/>
```

**SQL QUERY**

```
select * where { graph <http://rdf.geohistoricaldata.org/id/directories/photographies> {
    ?e a ont:Entry.
    ?e rdfs:label ?label.
    ?e prov:wasDerivedFrom ?directory.
    ?directory pav:createdOn 1862.
    ?e locn:address ?add.
    ?add locn:fullAddress ?fullAdd.
    ?add prov:wasGeneratedBy <http://rdf.geohistoricaldata.org/id/directories/activity/0001>.
    Filter regex(lcase(?label), "gallino").
}
```

**Filtres**

Propriétés

- Raison sociale: gallino
- Description: Ex: photo
- Adresse: Ex: rivoli

Tester une regex complexe.

Période

1861 1862

1790 1821 1853 1884 1915

Le filtre temporel permet de faire varier l'affichage des points préalablement chargés sur la carte sans lancer une nouvelle recherche.  
Données chargées pour la période 1861-1862.

## Questions de compétences

Combien y a-t-il de commerces de ce type localisés rue de Rivoli en 1856 ?

→ 14



```
PREFIX locn: <http://www.w3.org/ns/locn#>
PREFIX ont: <http://rdf.geohistoricaldata.org/def/directory#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX pav: <http://purl.org/pav/>
select (count(distinct ?e) as ?nombre) where {graph <http://rdf.geohistoricaldata.org/id/directories/photographies> {
?e a ont:Entry.
?e prov:wasDerivedFrom ?directory.
?directory pav:createdOn 1856.
?e locn:address ?add.
?add locn:thoroughfare ?voie.
?add prov:wasGeneratedBy <http://rdf.geohistoricaldata.org/id/directories/activity/0001>.
Filter regex(lcase(?voie), "rivoli").
}}
```

## Questions de compétences

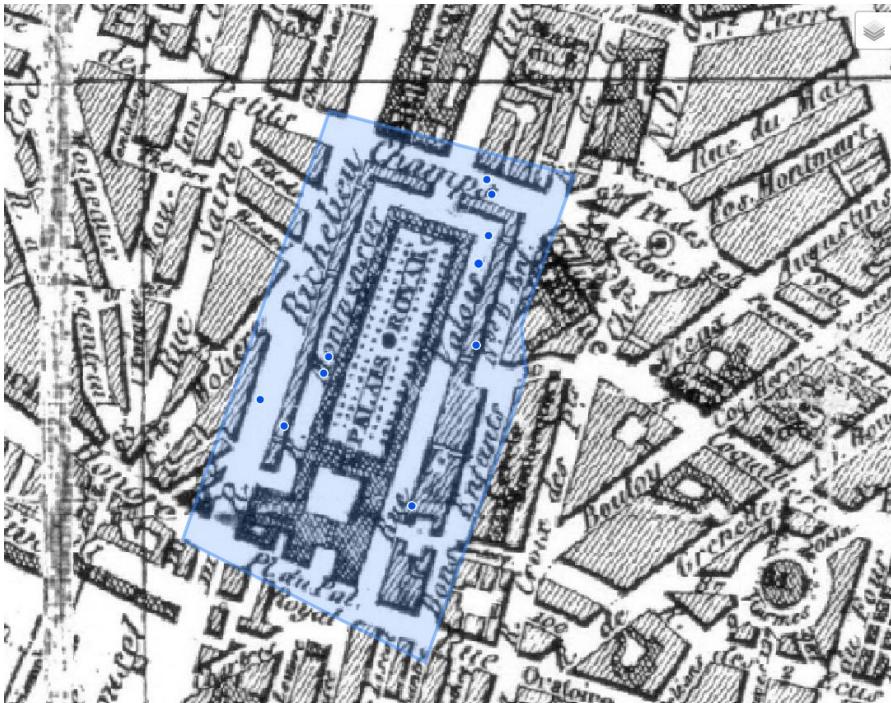
Quels sont les commerces situés dans une zone définie par un polygone ou un rectangle englobant en 1875 ?

```
PREFIX locn: <http://www.w3.org/ns/locn#>
PREFIX ont: <http://rdf.geohistoricaldata.org/def/directory#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX pav: <http://purl.org/pav/>
PREFIX gsp: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
select ?label where { graph <http://rdf.geohistoricaldata.org/id/directories/photographies> {
    ?e a ont:Entry.
    ?e rdfs:label ?label.
    ?e prov:wasDerivedFrom ?directory.
    ?directory pav:createdOn 1875.
    ?e locn:address ?add.
    ?add gsp:hasGeometry ?geom.
    ?geom gsp:asWKT ?wkt.
    FILTER (geof:sfIntersects(?wkt, "<http://www.opengis.net/def/crs/OGC/1.3/CRS84> Polygon((2.3360 48.8625,2.3360 48.8663,2.3385 48.8663,2.3385 48.8625,2.3360 48.8625))"^^gsp:wktLiteral))
}
}
```

## Questions de compétences

Quels sont les commerces situés dans une zone définie par un polygone ou un rectangle englobant en 1875 ?

Exemple de la zone située autour du Palais royal :



*Allévy*

*Barrès*

*Baur Vve*

*Bureau S*

*Crillon A*

*Derepas*

*Legros*

*Mustière*

*Prudent*

*Vantier*

## Questions de compétences

Quels commerces ont potentiellement déménagé au cours de la période 1860-1870 ?

```
PREFIX locn: <http://www.w3.org/ns/locn#>
PREFIX ont: <http://rdf.geohistoricaldata.org/def/directory#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX pav: <http://purl.org/pav/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
select distinct ?label ?voie ?labelo ?voieo where { graph <http://rdf.geohistoricaldata.org/id/directories/photographies> {
    ?e a ont:Entry.
    ?e rdfs:label ?label.
    ?e prov:wasDerivedFrom ?directory.
    ?directory pav:createdOn ?date.
    ?e locn:address ?add.
    ?add locn:thoroughfare ?voie.
    ?e owl:sameAs ?o.
    ?o rdfs:label ?labelo.
    ?o locn:address ?addo.
    ?addo locn:thoroughfare ?voieo.
    Filter ((?date > 1860) && (?date < 1870)).
    Filter (!sameTerm(?voie, ?voieo))
}}|
```



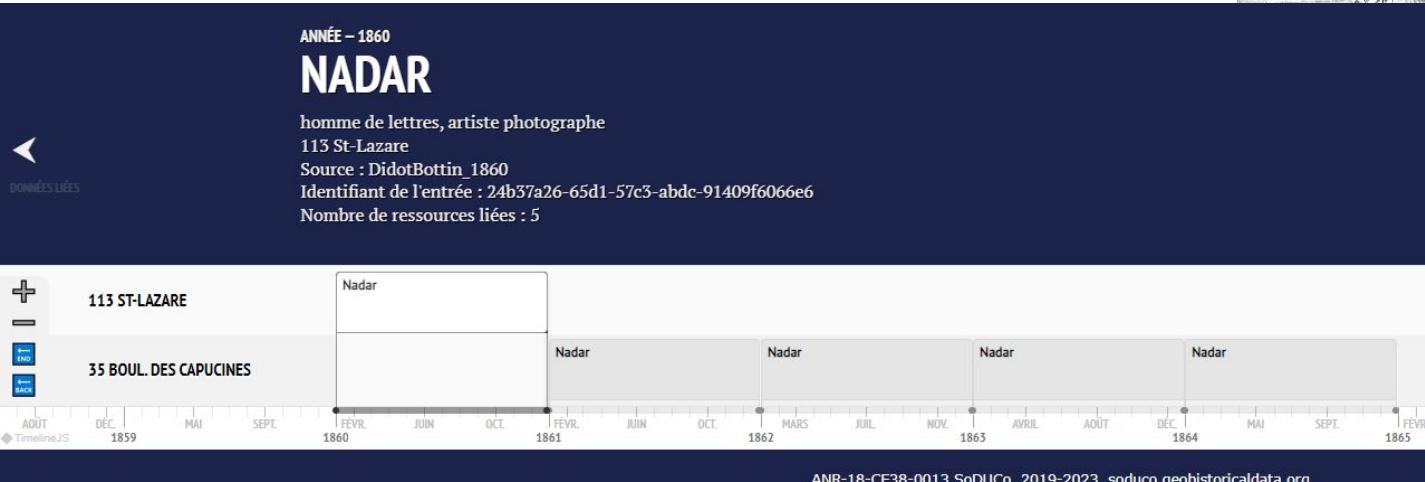
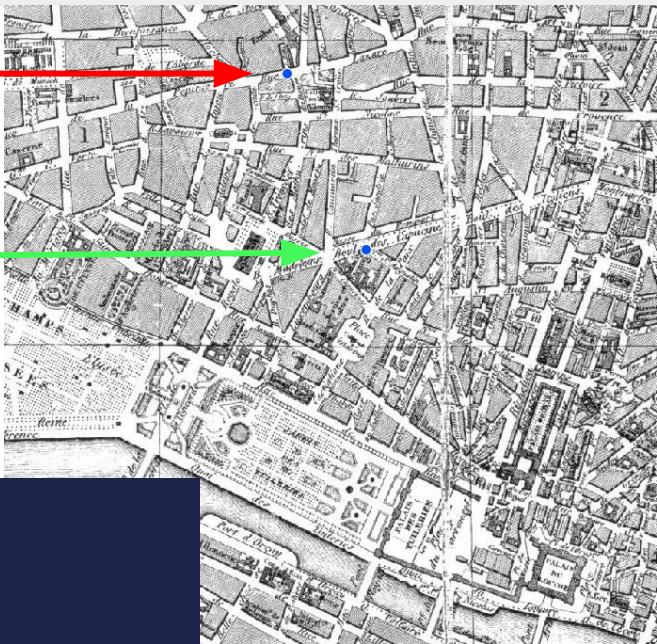
Exemple : Déménagement de Nadar de la Rue St-Lazare au Boulevard des Capucines vers 1860-1861

## Questions de compétences

St-Lazare

Quels commerces ont potentiellement déménagé au cours de la période 1860-1870 ?

Boulevard des Capucines



# Visualisation et évaluation des questions de compétences

## Questions de compétences

Quels commerces ont été repris par un autre commerçant exerçant la même activité ?

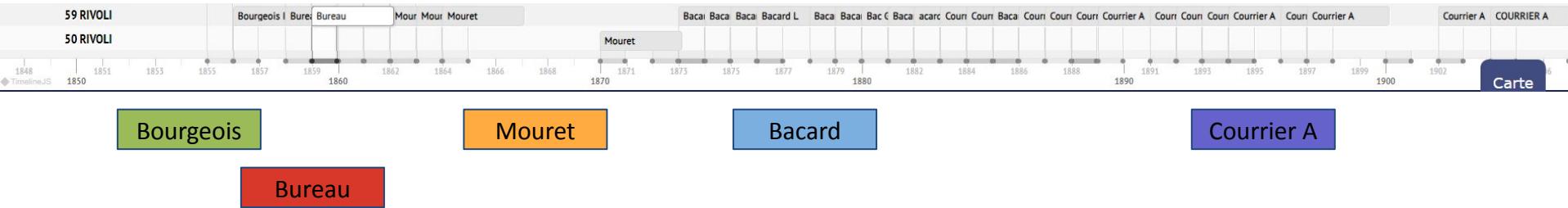
```
PREFIX locn: <http://www.w3.org/ns/locn#>
PREFIX ont: <http://rdf.geohistoricaldata.org/def/directory#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX prov: <http://www.w3.org/ns/prov#>
PREFIX pav: <http://purl.org/pav/>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
select distinct ?label ?fadd ?labelo ?faddo where { graph <http://rdf.geohistoricaldata.org/id/directories/photographies> {
    ?e a ont:Entry.
    ?e rdfs:label ?label.
    ?e locn:address ?add.
    ?add locn:fullAddress ?fadd.
    ?e owl:sameAs ?o.
    ?o rdfs:label ?labelo.
    ?o locn:address ?addo.
    ?addo locn:fullAddress ?faddo.
    Filter (!sameTerm(?label, ?labelo) && sameTerm(?fadd, ?faddo) )
}}
```

→ Exemple : Succession des photographes Bourgeois puis Bureau puis Mouret puis Bacard puis Courrier (A.) au 59 rue de Rivoli entre 1856 et 1908

# Visualisation et évaluation des questions de compétences

## Questions de compétences

Quels commerces ont été repris par un autre commerçant exerçant la même activité ?



Exemple : Succession des photographes Bourgeois puis Bureau puis Mouret puis Bacard puis Courrier (A.) au 59 rue de Rivoli entre 1856 et 1908

# Généralisation de l'approche avec d'autres professions

Liage réalisé pour 6 autres professions

Dataset	Nombre d'entrées	Nombre de liens
Architectes	91 706	1 621 050
Fabricants et marchands de chapeaux	46 680	604 817
Graveurs et marchands de cartes et plans	1 403	36 408
Magasins de nouveautés	33 562	465 455
Notaires	18 005	412 712
Vétérinaires	7024	155 905

→ Volumétries très variables

Evolutions différentes des professions au XIXème siècle

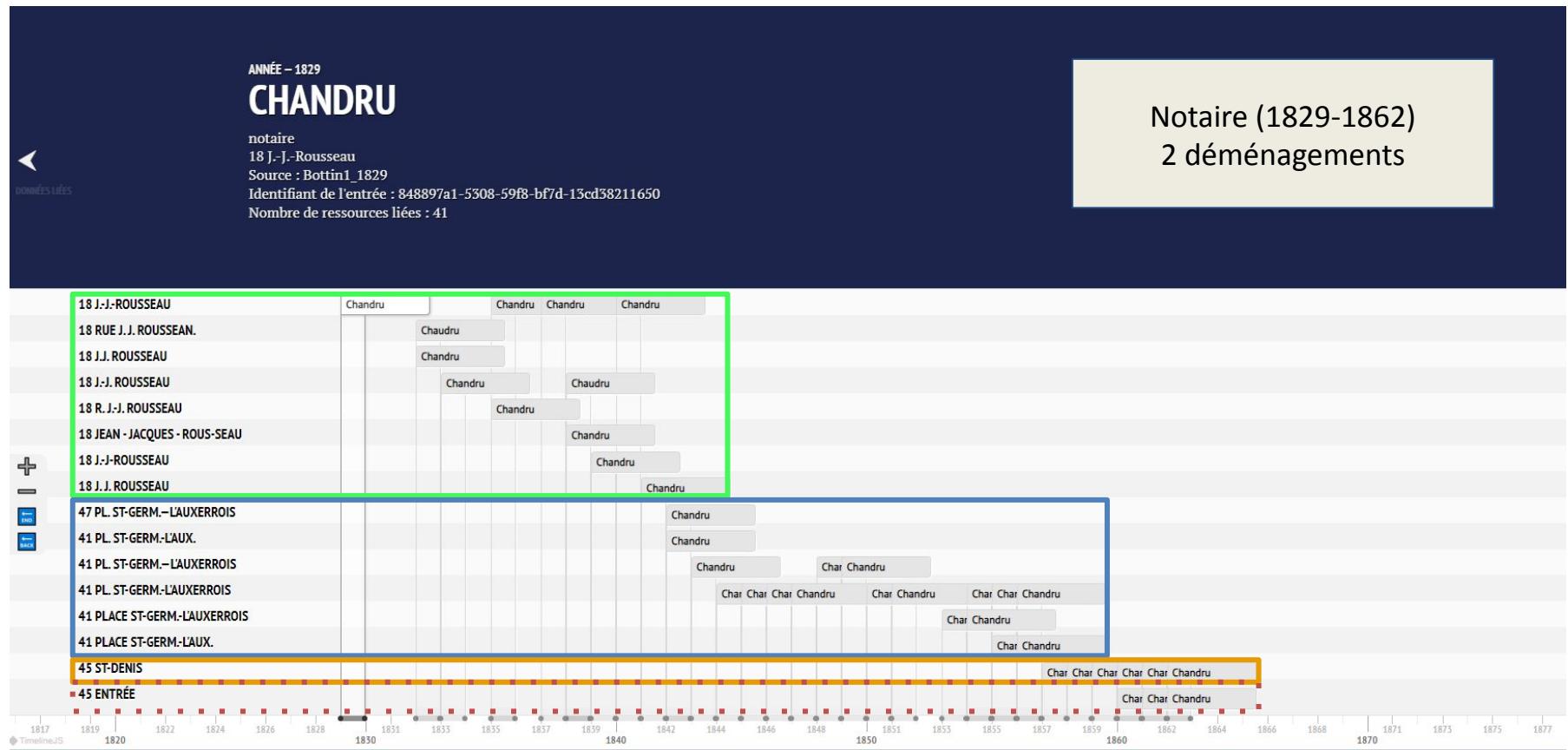
# Travaux en cours

## Exemple (Notaires) : Chandru



# Travaux en cours

## Exemple (Notaires) : Chandru



# Travaux en cours

## Exemple (Notaires) : Chandru

ANNÉE – 1864

### CHANDRU

ancien notaire

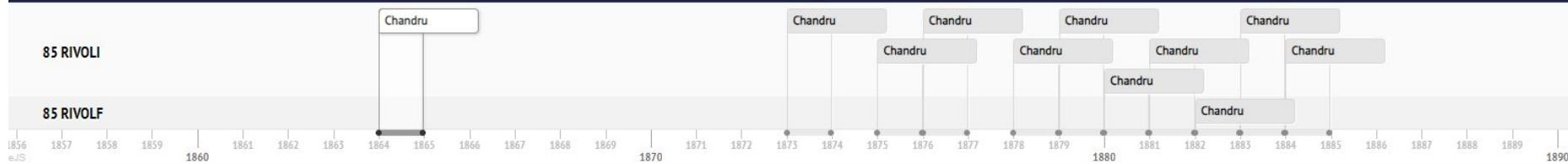
85 Rivoli

Source : DidotBottin\_1864

Identifiant de l'entrée : d662ba97-be23-5b0c-ad63-7458702407aa

Nombre de ressources liées : 11

Ancien notaire (1864)  
nouvelle adresse



# Travaux en cours

## Exemple (Notaires) : Chandru

Chandru

Adresse (annuaire) : 85 Rivoli  
Adresse (géocodage) : 16 Rue De Rivoli, Paris  
(Source: atlas\_jacoubet\_1836)  
Activité : ancien notaire  
Année de publication : 1864  
Annuaire : DidotBottin\_1864  
Identifiant de l'entrée : d662ba97-be23-5b0c-ad63-7458702407aa

16 Rivoli

24 Place Saint Germain  
L'Auxerrois

Chandru

Adresse (annuaire) : 41 pl. St-Germ.-l'Auxerrois  
Adresse (géocodage) : 24 Place Saint Germain  
L'Auxerrois, Paris (Source: atlas\_jacoubet\_1836)  
Activité : notaire  
Année de publication : 1851  
Annuaire : Didot\_1851  
Identifiant de l'entrée : 048f9491-2e50-537d-bbb2-305edcb80010



Atlas Municipal de Paris (1888). BHVP

45 Beauregard

Chandru

Adresse (annuaire) : 45 entrée  
Adresse (géocodage) : 45 Rue Beauregard, Paris  
(Source: atlas\_jacoubet\_1836)  
Activité : notaire  
Année de publication : 1860  
Annuaire : DidotBottin\_1860  
Identifiant de l'entrée : df0e828d-782a-5be0-b3df-92f4efd50114

18 J.-J.-Rousseau

85 Rivoli

45 St-Denis

## Liage avec d'autres ressources

The screenshot shows the FranceArchives portal interface. At the top left is the French Republic logo with the motto "Liberté, Égalité, Fraternité". The main header reads "FranceArchives PORTAIL NATIONAL DES ARCHIVES". On the right side of the header are links for "PAGES PROFESSIONNELLES" (with a person icon), "RECHERCHE SPARQL" (with a link icon), and "MENU" (with a three-line icon). Below the header is a search bar with a placeholder "Rechercher" and a magnifying glass icon. There are two checked checkboxes: "Archives référencées" and "Ressources du site". To the right of the search bar is a "Recherche avancée" link. At the bottom of the search bar area is a breadcrumb navigation: "Accueil > Archives nationales > Inventaire - MC/ET/III/1451 - MC/ET/III/1610, MC/RE/III/20 - ...".

### Minutes et répertoires du notaire Marcel CHANDRU, 28 novembre 1828 - 4 août 1862 (étude III)

Date : 1828 - 1862

Voir le contexte de la notice

#### Contenu :

Accès aux instruments de recherche de tous les notaires de l'étude III

#### Cote :

MC/ET/III/1451 - MC/ET/III/1610, MC/RE/III/20 - MC/RE/III/23

Tests de liage réalisés par Mathieu Zhral (Archives Nationales) avec l'extraction de 2022



Lancer la visite guidée

#### Où consulter le document :

Archives nationales

Voir sur le site d'origine

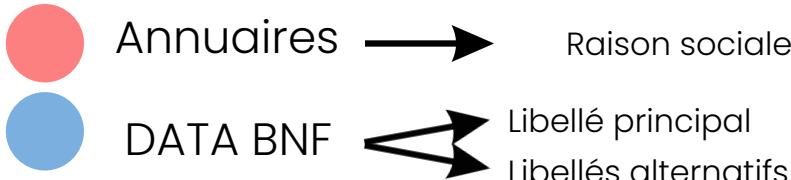
Coordonnées du service

Télécharger la notice en CSV

Télécharger en RDF

## Liage avec d'autres ressources

→ Test d'appariement par **comparaison numérique** des propriétés relatives au nom des photographes entre :

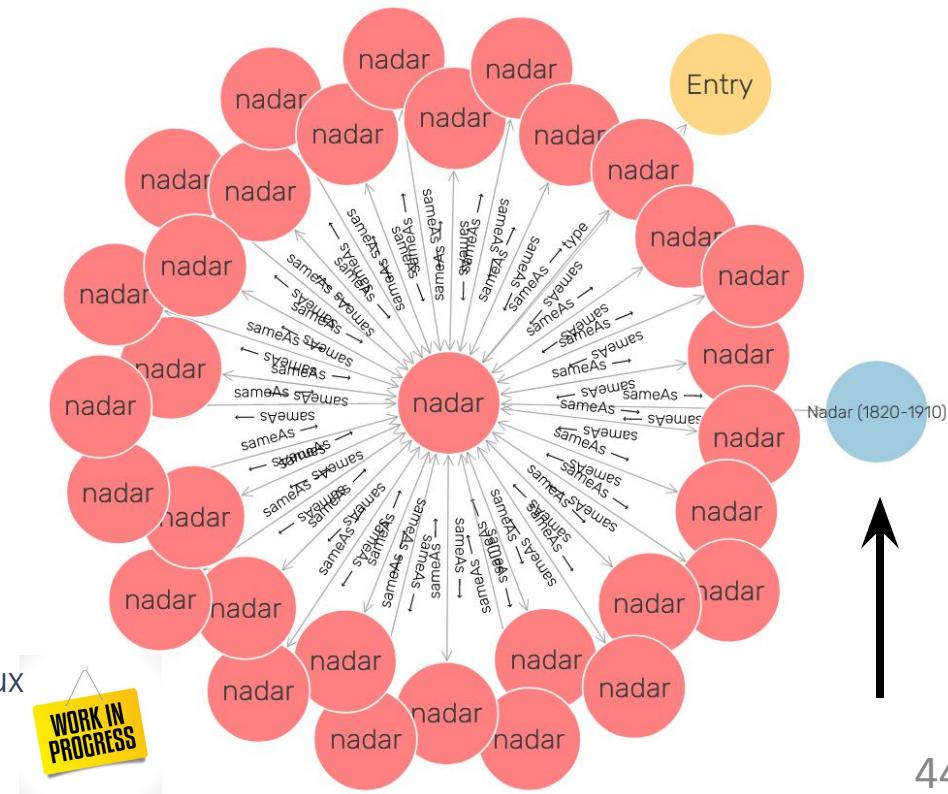


Nadar (1820-1910)

Pays : France  
Langue : Français  
Sexe : Masculin  
Naissance : Paris, 06-04-1820  
Mort : Paris, 21-03-1910  
Note : Photographe, journaliste, caricaturiste. - Aéronaute. - Son frère, Adrien Tournachon, peintre de formation et collaborant à son atelier de photographie, se servait du même pseudonyme  
Photographie  
Autres formes du nom : Gaspard-Félix Tournachon (1820-1910)  
Félix Tournachon (1820-1910)  
Félix Tournachon-Nadar (1820-1910)  
Félix Nadar (1820-1910)  
Nadar (1820-1910)  
Nadar père (1820-1910)  
ISNI : ISNI 0000 0001 2141 8535 (informations sur l'ISNI)

{BnF} Data

- 84 ressources "Photographes" sur data.bnf.fr
- 192 liens soit 10 ressources data.bnf.fr liées aux ressources issues des Annuaires



## Conclusion

→ Démarche fonctionnelle d'exploitation des entrées issues des annuaires

- extraction fine et en masse
- entrées structurées, souvent bruitées.

→ Travaux à venir :

- évaluation quantitative des appariements
- liage avec des ressources externes
- aide pour l'explicitation des relations spatio-temporelles entre les entrées (succession, déménagement...) par les experts du domaine
- adaptation de la méthode à d'autres types de données historiques

→ Méthode, code, application et données accessibles sur le web.

# Merci pour votre attention

**Tutoriel, ontologie, ressources, code et questions de compétences**

[https://github.com/soduco/atelier\\_graphes\\_geohistoriques\\_annuaires](https://github.com/soduco/atelier_graphes_geohistoriques_annuaires)

## Application

[bit.ly/graphe\\_geohistorique\\_soduco](https://bit.ly/graphe_geohistorique_soduco)

## Endpoint

<https://dir.geohistoricaldata.org/sparql>

Solenn Tual, Nathalie Abadie, Bertrand Duménieu, Joseph Chazalon et Edwin Carlinet. Création d'un graphe de connaissances géohistorique à partir d'annuaires du commerce parisien du 19ème siècle: application aux métiers de la photographie. *IC 2023, 34èmes journées francophones d'Ingénierie des connaissances*, Strasbourg, France, 3-7 Juillet 2023. [hal-04121643](https://hal.archives-ouvertes.fr/hal-04121643)

[solenn.tual@ign.fr](mailto:solenn.tual@ign.fr) [nathalie-f.abadie@ign.fr](mailto:nathalie-f.abadie@ign.fr) [bertrand.dumenieu@ehess.fr](mailto:bertrand.dumenieu@ehess.fr) [joseph.chazalon@epita.fr](mailto:joseph.chazalon@epita.fr) [edwin.carlinet@epita.fr](mailto:edwin.carlinet@epita.fr)

## Références

Nathalie Abadie, Edwin Carlinet, Joseph Chazalon, et Bertrand Duménieu. **A Benchmark of Named Entity Recognition Approaches in Historical Documents Application to 19th Century French Directories.** Number 13237 in Document Analysis Systems. DAS 2022., La Rochelle, France, May 2022. Springer, Cham.

Bertrand Duménieu. **Localiser les adresses anciennes ? Plans historiques & annuaires parisiens au XIXe siècle.** Journée d'étude scientifique "Pour un traitement scientifique automatisé des cartes anciennes", Festival Printemps des Cartes 2023

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.

Danai Symeonidou, Vincent Armant, Nathalie Pernelle, and Fatiha Saïs. **SAKey : Scalable Almost Key Discovery in RDF Data.** In proceedings of the 13th International Semantic Web Conference, ISWC 2014, volume Lecture Notes in Computer Science of The Semantic Web – ISWC 2014, pages 33–49, Riva del Garda, Italy, October 2014. Springer.

**Silk Linked Data Integration Framework : outil de liage de données,** <https://github.com/silk-framework/silk>

**Pero OCR,** <https://pero-ocr.fit.vutbr.cz/>