

SoDUCo : croisement de sources géo-historiques pour l'étude de l'évolution de Paris de 1789 à 1950

Séminaire IXXI "Extraction, traitement et visualisation de données complexes en géographie (XVIIIe s. - XIXe s.)" - 20/02/2023

Nathalie Abadie¹, Stéphane Baciocchi², Charly Bernard¹, Edwin Carlinet³, Paul Chapron¹, Joseph Chazalon³, Yizi Chen^{1,3}, Pascal Cristofoli², Bertrand Duménieu^{1,2}, Mathieu Fernandez², Julie Gravier², Melvin Hersent¹, Clément Mallet¹, Julien Perret^{1,2}, Juste Raimbault¹, Solenn Tual^{1,3}



Social Dynamics in Urban Context

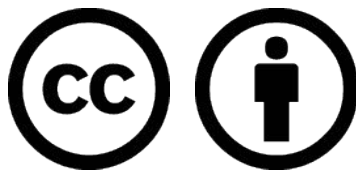


1. LASTIG, Univ Gustave Eiffel, ENSG/IGN
2. LaDéHiS, CRH, EHESS
3. LRE, EPITA



(Ré)utiliser cette présentation

Licence **Creative Common BY 4.0**



N'hésitez pas à **réutiliser** ces contenus et nous **citer**.
*Sauf citation ou mention contraire,
nous sommes les auteurs des contenus.*

Consultez le diaporama pendant et après la présentation

<https://bit.ly/20230220-SODUCO-IXXI>



Partie 1 : SoDUCo

Projet ANR **SoDUCo**

(Social Dynamics in Urban Context)

Open tools, models, and data - Paris and its suburbs, 1789-1950

<https://soduco.github.io/>

<https://anr.fr/Projet-ANR-18-CE38-0013>

2019-2023



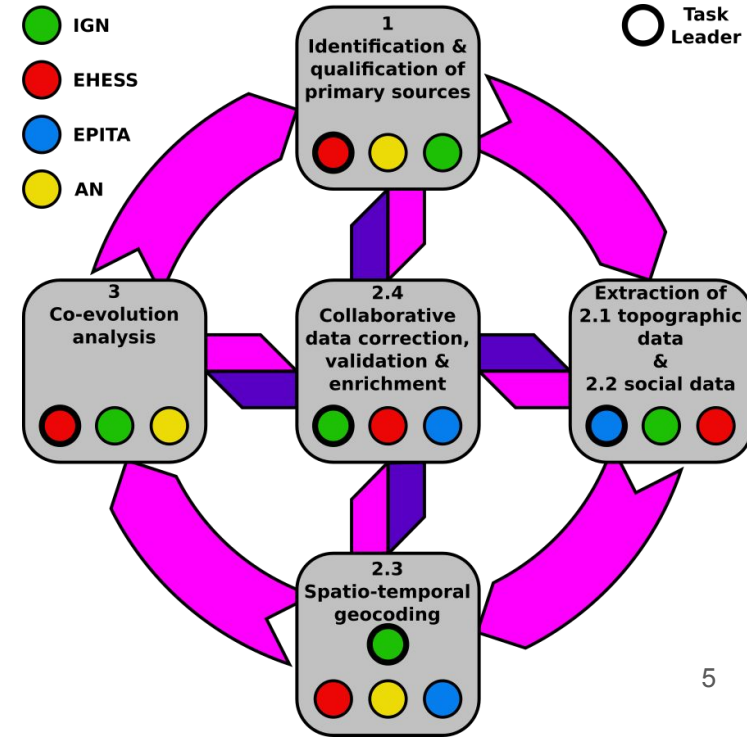
Objectifs

Approche globale et ouverte pour l'analyse des *évolutions*

Analyse critique des *sources primaires*

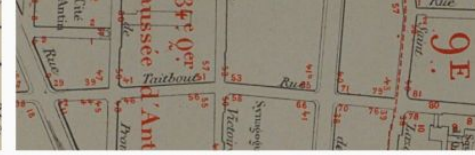
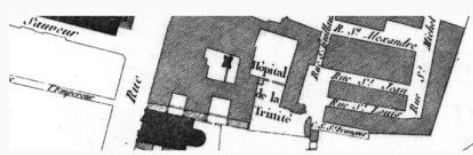
Co-construction des *données géo-historiques*

Paris et sa banlieue de 1789 à 1950



Les **sources**, leur qualification et leur diffusion

Sources primaires



**GRAVEURS EN GÉOGRAPHIE
ET TOPOGRAPHIE.**
Arnoult, *Serpente* 16.
Berthe, cit. 1827, *Noyers* 46.

Popincourt 150.
Dehaeck, *Charenton* 166.
Delaigle, p. *St-Germain-des-Prés* 6.
Delangle, *Clotaire* 3.
Delaporte, *Cassette* 12.

MERCERIES, SOIERIES ET NOUVEAUTÉS (MARCHANDS DE).

Achard, passage du Panorama, 61.
Alavoine, de soieries et nouveautés pour
chaussure, rue de la Poterie - des
Halles, 37.

Belin, r. Saint-Martin, 125.
Bellard et Lhulier (Mesd.), nouveautés,
r. de la Paix, 22.

Identification des sources pertinentes

- Sources cartographiques
- Annuaire professionnels

Qualification des sources (contenu, imperfections, etc.)

Publication et diffusion des sources

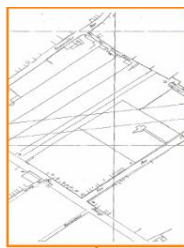
Sources primaires cartographiques



Atlas de Verniquet



Cadastré



Atlas de Jacoubet



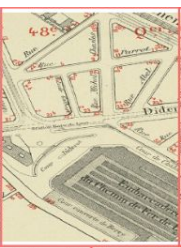
Atlas Mun. 1882



Atlas Mun. 1893



Atlas Mun. 1900



Atlas Mun. 1912



Atlas Mun. 1926



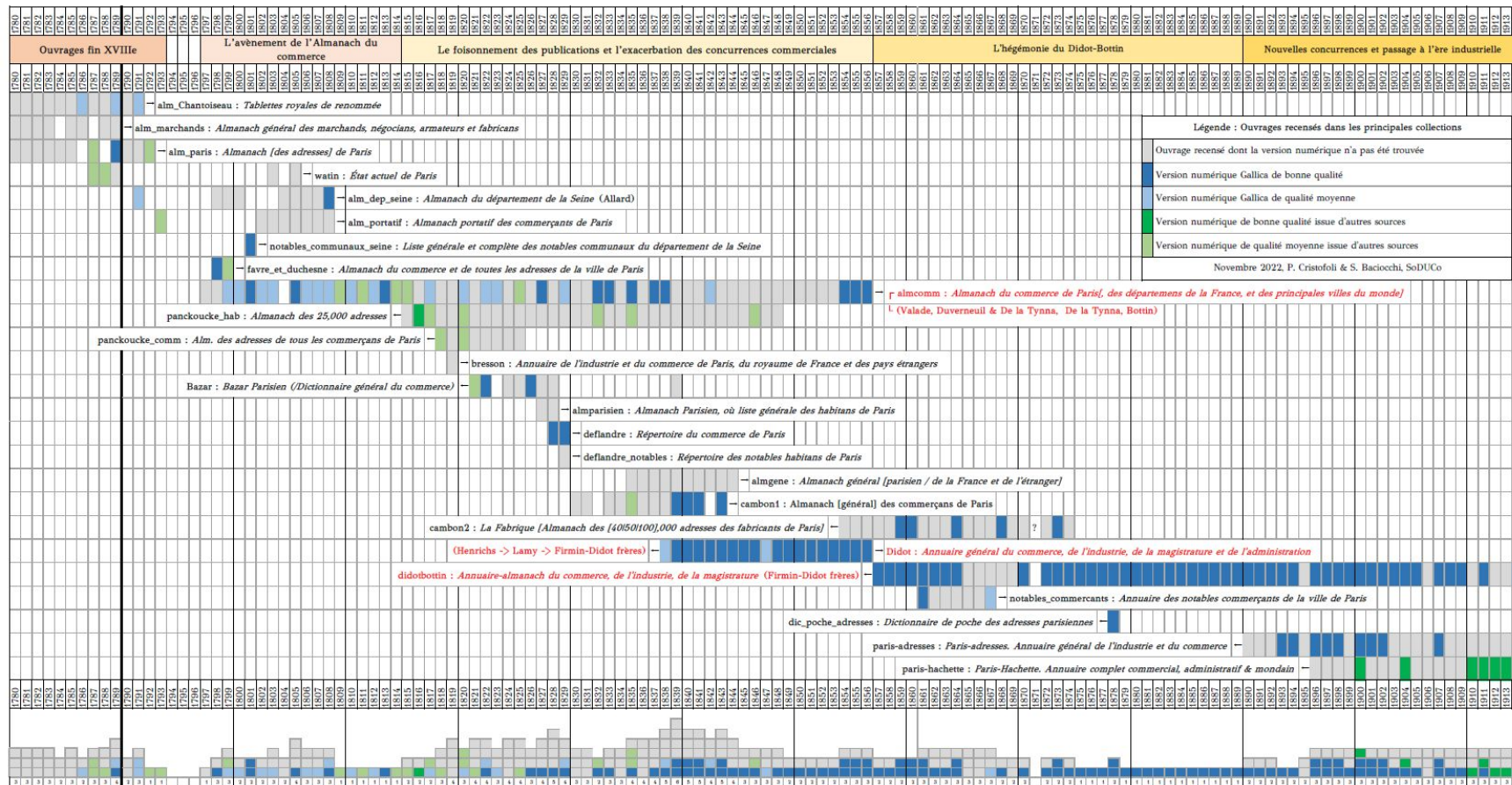
Atlas Mun. 1937



Short title	Spatial footprint	Temporal footprint	Volume (of sheets)	Scanned	Georeferenced	State of vectorisation	Streets	City blocks	Parcels	Street numbers
Atlas de Verniquet		1789-1799	72	●	●	■■■■	●	●	○	○
Minutes de l'atlas de Verniquet	City boundary	1783-1789	~ 1000	●	○	■■■■	●	●	○	○
Plan de Maire	& adjacent suburbs	1808	20	●	○	■■■■	●	●	○	○
Atlas de Jacoubet		1827-1836	54	●	●	■■■■	●	●	○	○
Carte de Paris et de ses fortifications		1845-1849	1	●	○	■■■■	●	●	○	○
Cadastré de Vasserot ¹	City boundary bef. 1860	1810-1836	910	●	○	■■■■	●	●	○	○
Vasserot&Bellanger cadastré	Paris right bank bef. 1860	1830-1850	155	●	○	■■■■	●	●	○	○
Municipal atlas		1868	16	●	●	■■■■	●	●	○	○
Mun. atlas & roadworks		1871	1	●	○	■■■■	●	●	○	○
Reconstructed city cadastré		1880-1910	1	●	○	■■■■	●	●	○	○
Lanée map		1879	1	●	○	■■■■	●	●	○	○
Municipal atlas updated		1888	16	●	●	■■■■	●	●	○	○
Hachette map		1900	1	●	○	■■■■	●	●	○	○
Municipal atlas updated	Paris boundary after 1860	1905	16	●	○	■■■■	●	●	○	○
Taride map		1907	1	●	○	■■■■	●	●	○	○
Municipal atlas updated		1920	16	●	○	■■■■	●	●	○	○
Guilmin map		1926	1	●	○	■■■■	●	●	○	○
Municipal atlas updated		1929	16	●	○	■■■■	●	●	○	○
Dufrénoy map		1937	1	○	○	■■■■	●	●	○	○
Municipal atlas of Paris		1948	23	○	○	■■■■	●	●	○	○
City cadastré		1999	n.c.	●	●	■■■■	●	●	○	○
Napoleonic cadastré	Adj. suburbs bef. 1860	1808-1825	n.c.	●	○	■■■■	○	○	○	○
Napoleonic cadastré updated		1830-1850	n.c.	●	○	■■■■	○	○	○	○

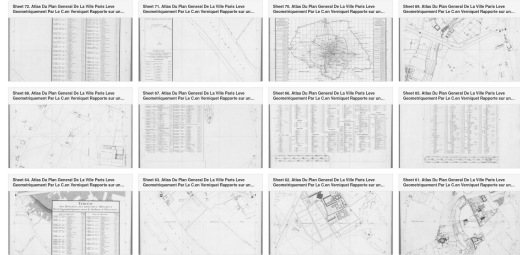
Inventaire des sources contenant des listes d'adresses

Inventaire annuel par collections d'ouvrages avec mention des versions numériques librement disponibles



Un catalogue pour la diffusion des sources

<https://catalog.geohistoricaldata.org/>



Un catalogue spécialisé pour des ressources archivistiques et des données géographiques produites par un projet de recherche SHS/SIG :

- **cartes anciennes** et leurs **géoréférences**;
- **sources textuelles numérisées** : annuaires du commerce de Paris XIXe;
- **résultats d'extraction** : données socio-professionnelles géocodées, etc.

Une plateforme ouverte pour **rassembler et exposer** des ressources éparpillées dans diverses institutions patrimoniales pour former des **corpus cohérents**

Stocker, diffuser et documenter les productions (*ouvertes*) du projet, en particulier les **éditions critiques et géoréférences des plans de Paris et les numérisations des annuaires du commerce**

L'extraction d'information dans les sources

Extraction d'information dans les sources **cartographiques**

Un problème **complexe**



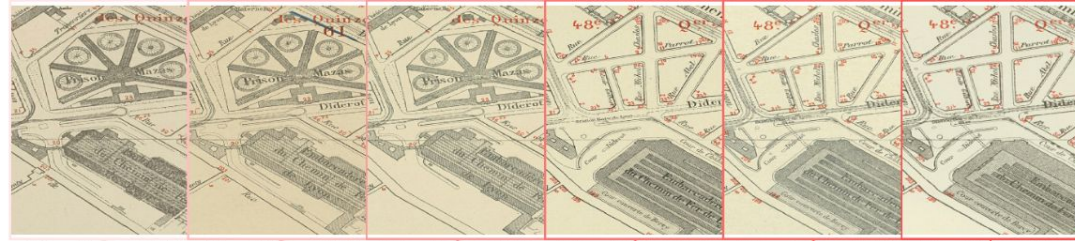
Atlas de Verniquet



Cadastre



Atlas de Jacoubet



Atlas Mun. 1882 Atlas Mun. 1893 Atlas Mun. 1900 Atlas Mun. 1912 Atlas Mun. 1926 Atlas Mun. 1937



Grande **variété** (représentation et contenu)

Des **imperfections** difficiles à prendre en compte

Compromis entre **généricité** et **spécificité** des approches

Choix d'un cas particulier intéressant : la série des **atlas municipaux** ~1867-1937

Cohérence de représentation et de contenu (et de système de coordonnées)

Redondance & suivi des **évolutions** morphologiques de Paris

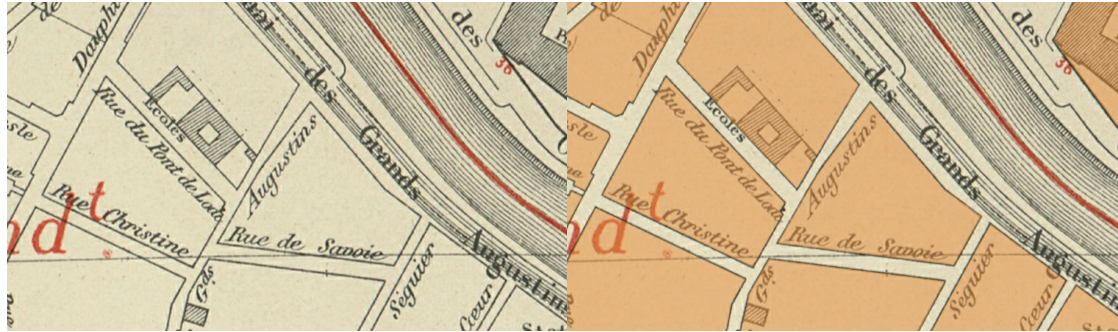
Les atlas municipaux de Paris

Un corpus de documents numérisés de la Bibliothèque de l'Hôtel de Ville de Paris (BHdV) :

1. Plan général de la ville de Paris et de ses environs comprenant les bois de Boulogne et de Vincennes ... 1866. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <https://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935006>
2. Atlas administratif des 20 arrondissements de la ville de Paris ... 1868. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000261673>
3. Atlas administratif des 20 arrondissements de la ville de Paris... 1870. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935043>
4. Atlas administratif des 20 arrondissements de la ville de Paris... 1872. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935076>
5. Atlas municipal des vingt arrondissements de la ville de Paris... 1876. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935091>
6. Atlas municipal des vingt arrondissements de la ville de Paris... 1878. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935100>
7. Atlas municipal des vingt arrondissements de la ville de Paris... 1880. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935105>
8. Atlas municipal des vingt arrondissements de la ville de Paris... 1882. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935108>
9. Atlas municipal des vingt arrondissements de la ville de Paris... 1885. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935111>
10. Atlas municipal des vingt arrondissements de la ville de Paris... 1886. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935114>
11. Atlas municipal des vingt arrondissements de la ville de Paris... 1887. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935115>
12. Atlas municipal des vingt arrondissements de la ville de Paris... 1849 (1888). Bibliothèque de l'Hôtel de Ville. Ville de Paris. <https://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935116/>
13. Atlas municipal des vingt arrondissements de la ville de Paris... 1893. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935117>
14. Atlas municipal des vingt arrondissements de la ville de Paris... 1894. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935118>
15. Atlas municipal des vingt arrondissements de la ville de Paris... 1895. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935508>
16. Atlas municipal des vingt arrondissements de la ville de Paris... 1898. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935509>
17. Atlas municipal des vingt arrondissements de la ville de Paris... 1900. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935510>
18. Atlas municipal des vingt arrondissements de la ville de Paris... 1905. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935514>
19. Atlas municipal des vingt arrondissements de la ville de Paris... 1909. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935515>
20. Atlas municipal des vingt arrondissements de la ville de Paris... 1912. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0002055151>
21. Atlas municipal des vingt arrondissements de la ville de Paris... 1912. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0002055156>
22. Atlas municipal des vingt arrondissements de Paris... 1925. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935524>
23. Atlas municipal des vingt arrondissements de Paris... Edition révisée en 1928 et 1929... 1er juin 1929. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935525>
24. Atlas municipal des vingt arrondissements de Paris... 1929-1936. Bibliothèque de l'Hôtel de Ville. Ville de Paris. <http://bibliotheques-specialisees.paris.fr/ark:/73873/pf0000935528>

Un sous-ensemble de problèmes

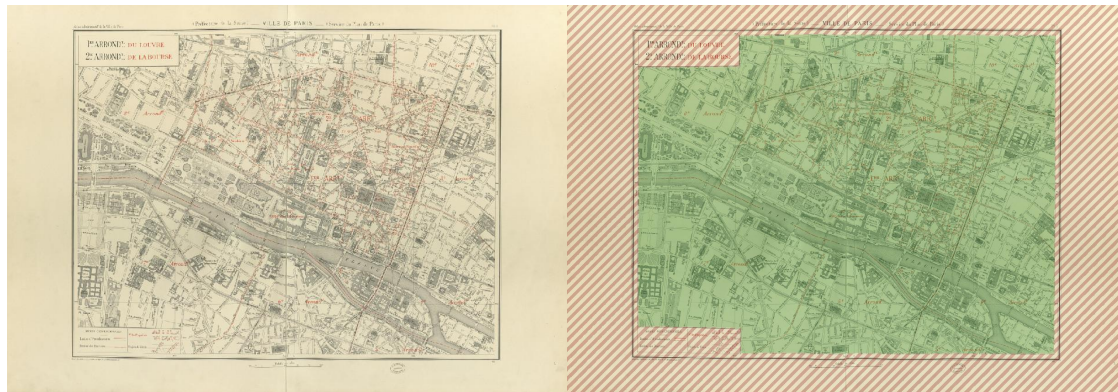
Détection des îlots morphologiques



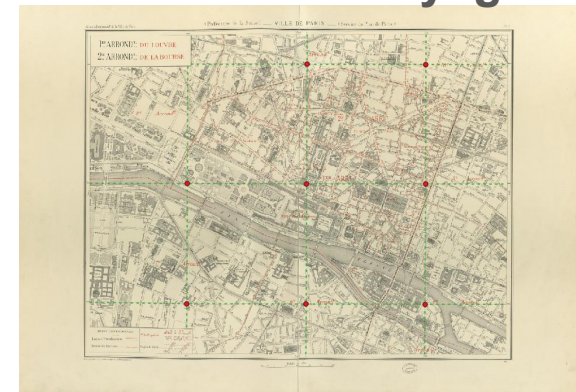
[ICDAR 2021 Competition on Historical Map Segmentation](#)

[ICDAR 2021 Competition on Historical Map Segmentation — Dataset | Zenodo](#)

Détection du contenu de la carte



Détection du carroyage

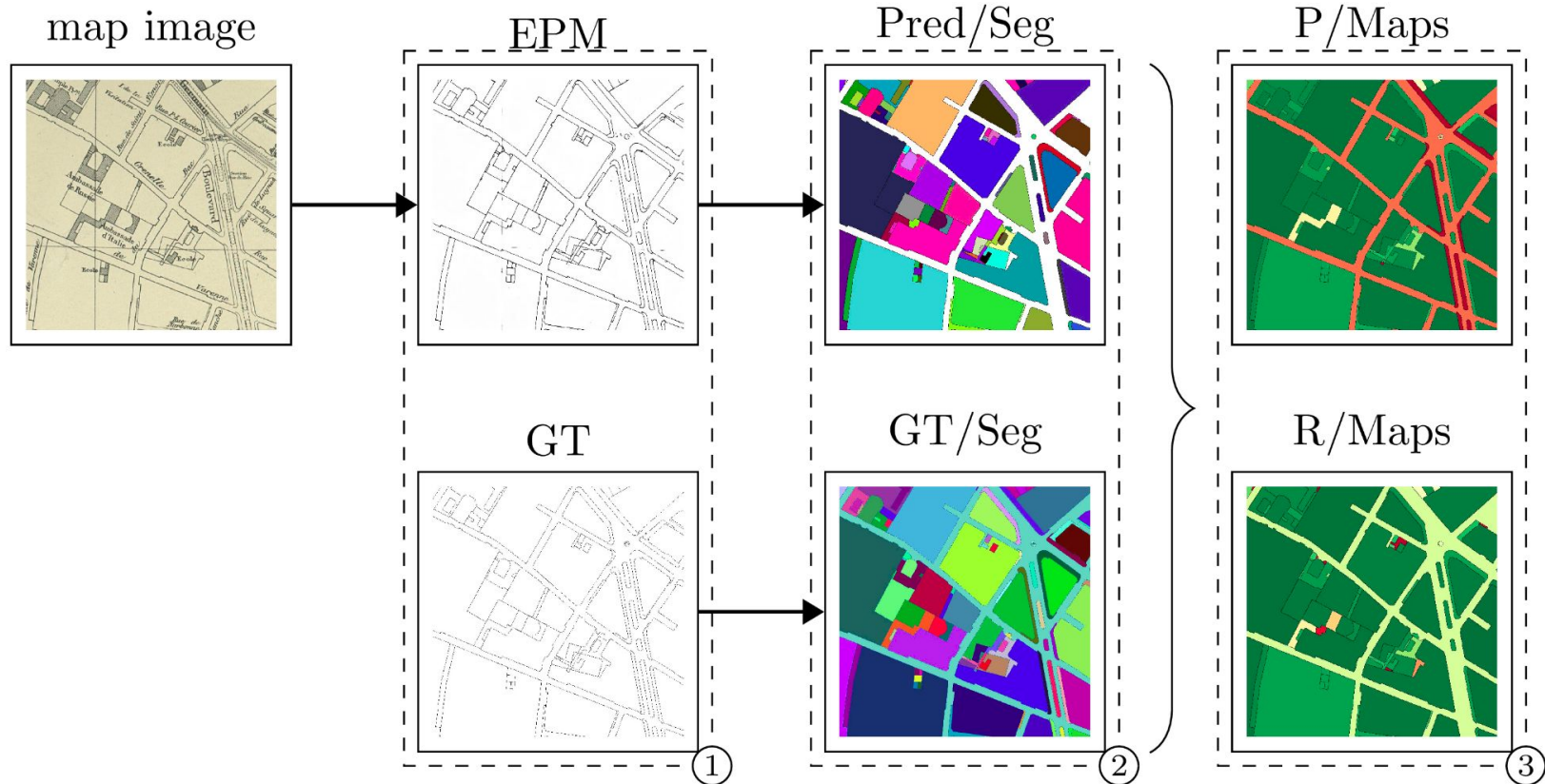


Extraction des bâtiments et des îlots dans les cartes historiques



Top: Historical map images; bottom: buildings and building blocks.

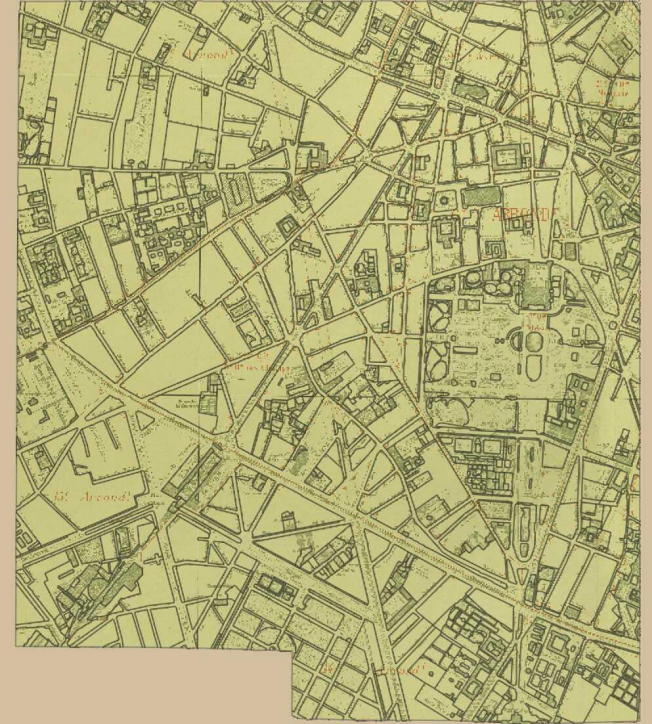
Une chaîne de traitement pour l'extraction de polygones



Résultat de vectorisation d'une feuille de l'Atlas Municipal

Vectorisation automatique pour l'aide à la saisie

Reprise collaborative



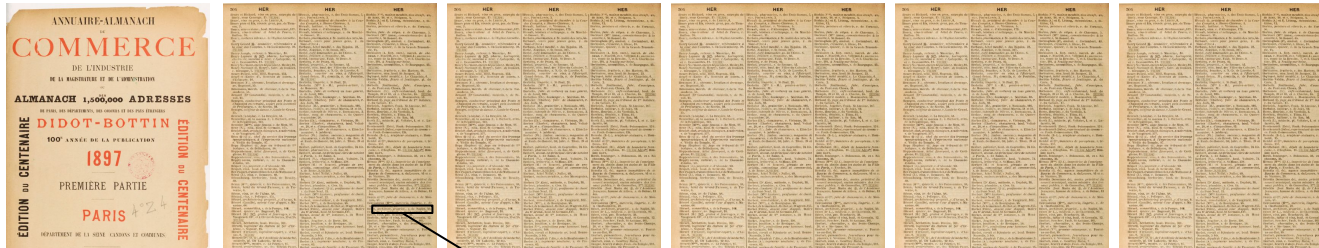
Des outils collaboratifs de vectorisation / validation

<https://workbench.geohistoricaldata.org/>



Extraction d'information dans les sources **textuelles**

Des documents numérisés aux données structurées



Pages d'annuaires numérisées

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

Identification des entrées

Entrée d'annuaire

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

Transcription du texte

Texte numérique de l'entrée

PERSONNE ACTIVITÉ RUE NUMÉRO RUE NUMÉRO
Lemonnyer, plomberie, r. de Bondy, 86, et r. Bouchardon, 1.

Annotation du texte

Entités nommées

Structuration des entrées

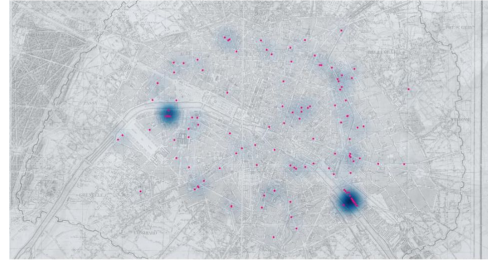
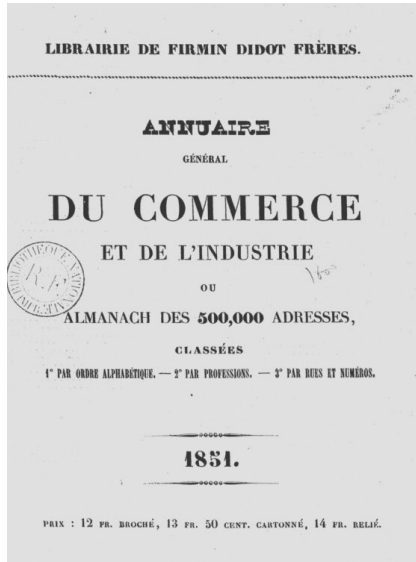
Joseph va vous en parler en détail !



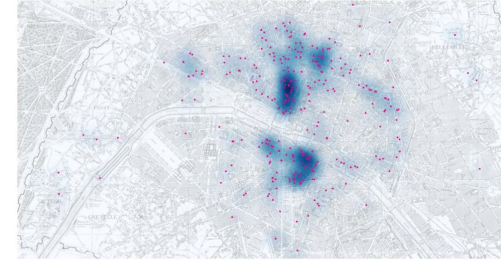
Base de données intégrée

Le Géocodage historique

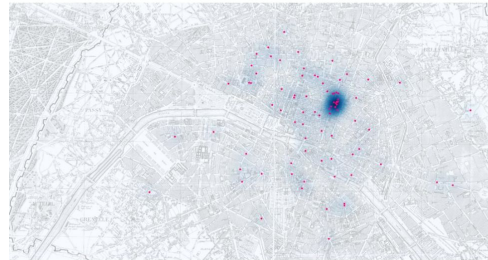
Le Géocodage historique



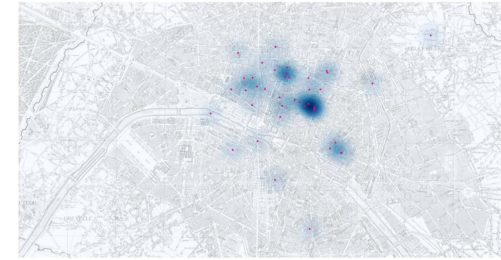
Métiers du bois



Avocats



Fabriquants de parapluies



Gantiers

Ravrio et comp., *fabr. de bronzes et curiosités*, r. Richelieu, 93; la fabrique rue Montmartre, 161.

OCR

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

NER

Géocodage historique

Géocoder les annuaires et dégager l'index géographique des adresses de Paris

- ❑ Le **géocodage** est une opération qui convertit une **référence spatiale indirecte** en une **référence spatiale directe**

55 rue Meslay

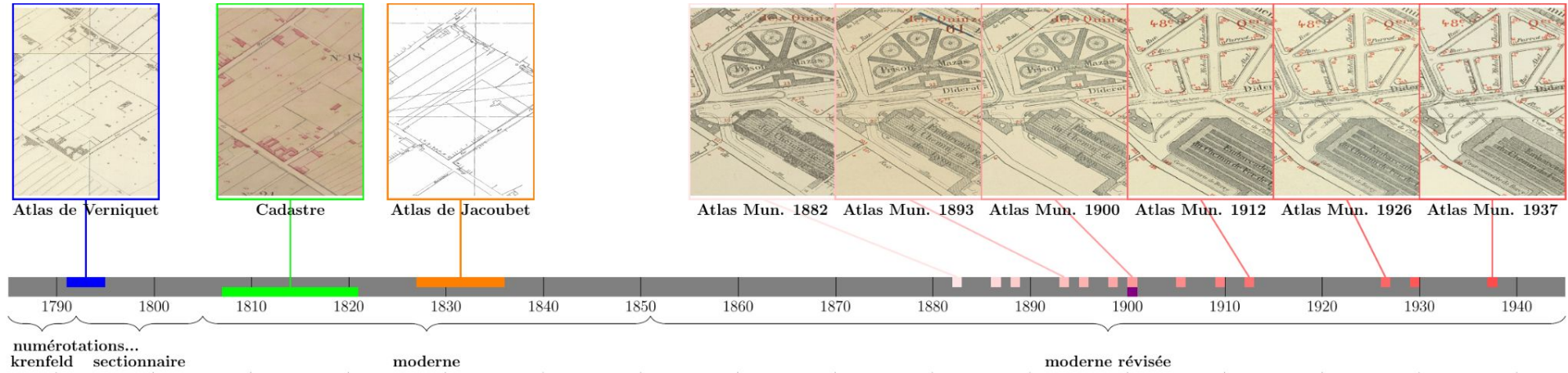


48° 52' 55.37" N , 2° 21' 11.62" E

- ❑ Bon nombre d'adresses anciennes ayant disparu, les **plans numérotés de Paris** servent à ancrer les adresses extraites des annuaires.



Des système d'adressage qui évoluent



Les grands atlas géoréférencés pour géocoder les adresses

Atlas général de la ville, des faubourgs et des monuments de Paris,
par Th. Jacoubet, architecte. Gravé par Bonnet. 🕒 1827 - 1836



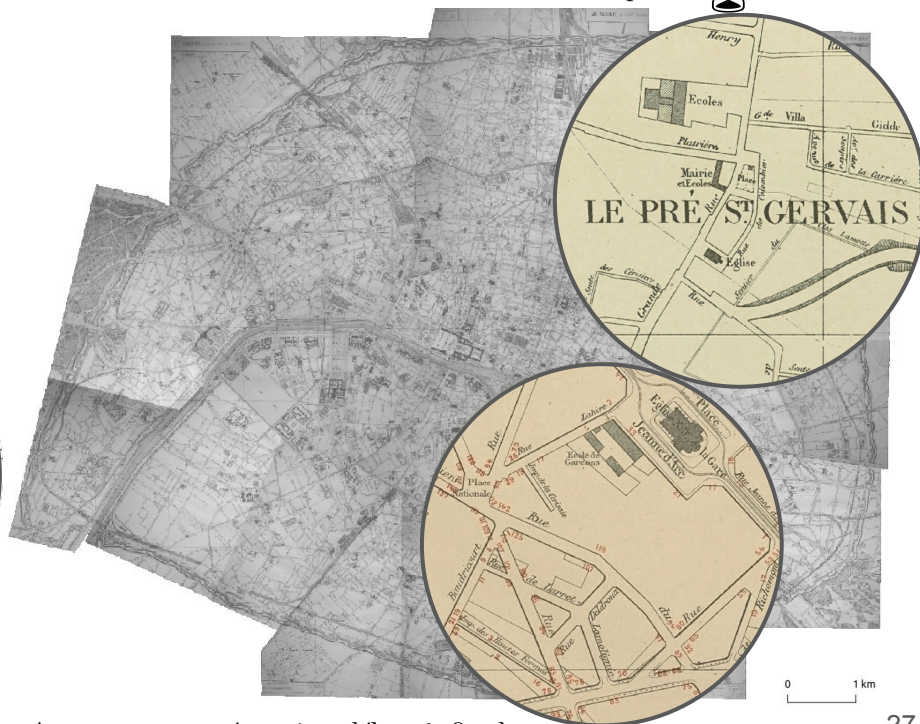
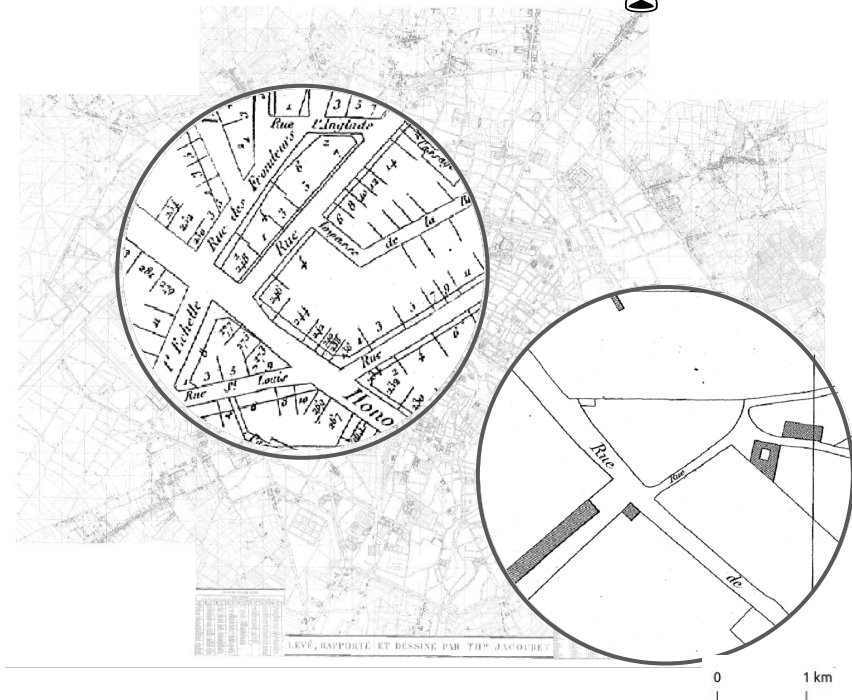
Atlas municipal des vingt arrondissements de la Ville de Paris.
Dressé sous l'administration de M. Poubelle, préfet. 🕒 1886 - 1888



Les grands atlas géoréférencés pour géocoder les adresses

Atlas général de la ville, des faubourgs et des monuments de Paris,
par Th. Jacoubet, architecte. Gravé par Bonnet. ⌚ 1827 - 1836

Atlas municipal des vingt arrondissements de la Ville de Paris.
Dressé sous l'administration de M. Poubelle, préfet. ⌚ 1886 - 1888



Au centre : amorces du bâti numérotées
Aux marges : parcellaire rural

Au centre : numérotation début & fin de tronçon
Aux marges : tables d'assemblage cadastral

Base de données géohistorique support du géocodage

Deux types d'entités géohistoriques vectorisés dans les plans : les **points adresse** et le **réseau des rues**

Atlas Jacoubet : réseau viaire et adresses dérivées des données Vasserot-ALPAGE⁽¹⁾
30 000 adresses, 1395 rues nommées



Atlas municipal : réseau viaire; adresses construites par interpolation
100 000 adresses, 3684 rues nommées



Les données vectorisées sont structurées en quadruplets

{ **dénomination, géométrie, temps valide, plan source** }

Un géocodeur historique sensible aux temps valides

Adresse
référence spatiale indirecte
r. St. Landry Cité, 2

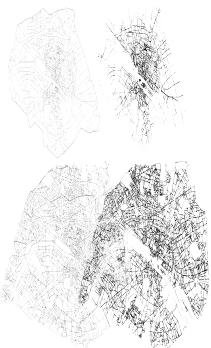
Année cible
1838




Géocodeur



Quadruplets
rues & adresses

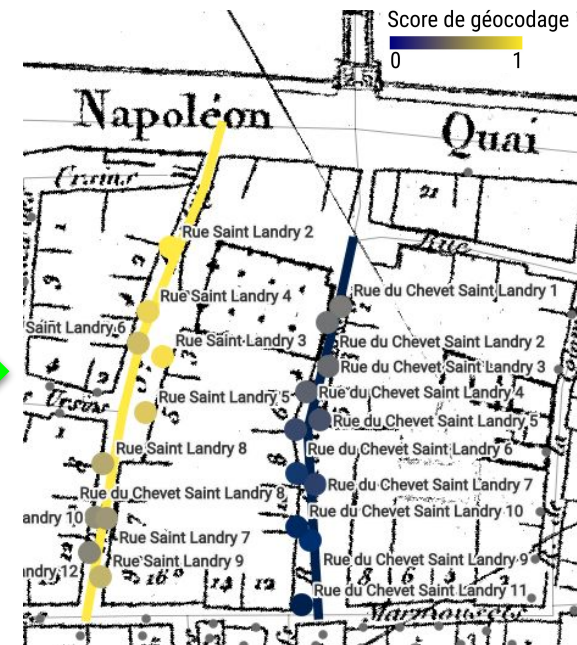
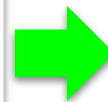


1.  Analyse syntaxique (*parsing*)

< rue > < quartier > < numéro >
r. St. Landry Cité, 2

2.  Recherche documentaire (*collect, filter & score*)

3.  Tri temporel (*reranking*)



Les annuaires, une source de références spatiales indirectes, fines, en masse...

Quelques ordres de grandeur : 113 annuaires extraits, 10 millions d'entrées détectées, environ autant d'entités **LOC** et d'entités **CARDINAL**

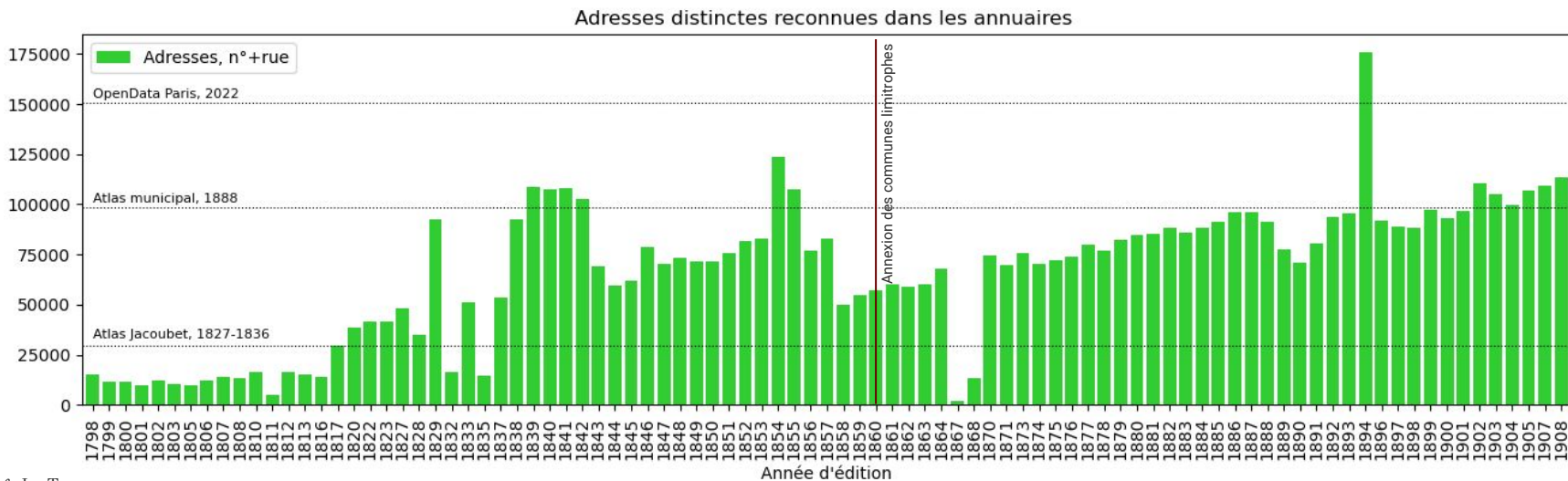


Les entités **LOC** et **CARDINAL** se combinent en **adresses à deux niveaux de détail** :

1. la rue : “Meslay”
2. le numéro adresse : “55 Meslay” et “57 Meslay”

... et spatio-temporelles

Le caractère sériel des annuaires permet de suivre une adresse dans la durée et d'identifier les **transitions** entre adresses.



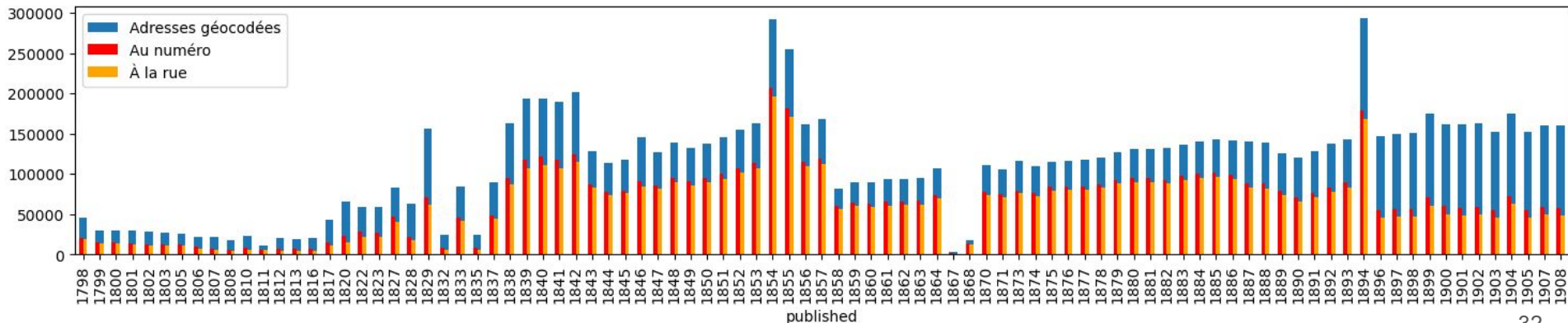
Duverneuill & La Tynna
La Tynna
Bottin
Didot, Didot-Bottin

Résultats du géocodage des annuaires

Environ 2/3 des adresses ont été géocodées :

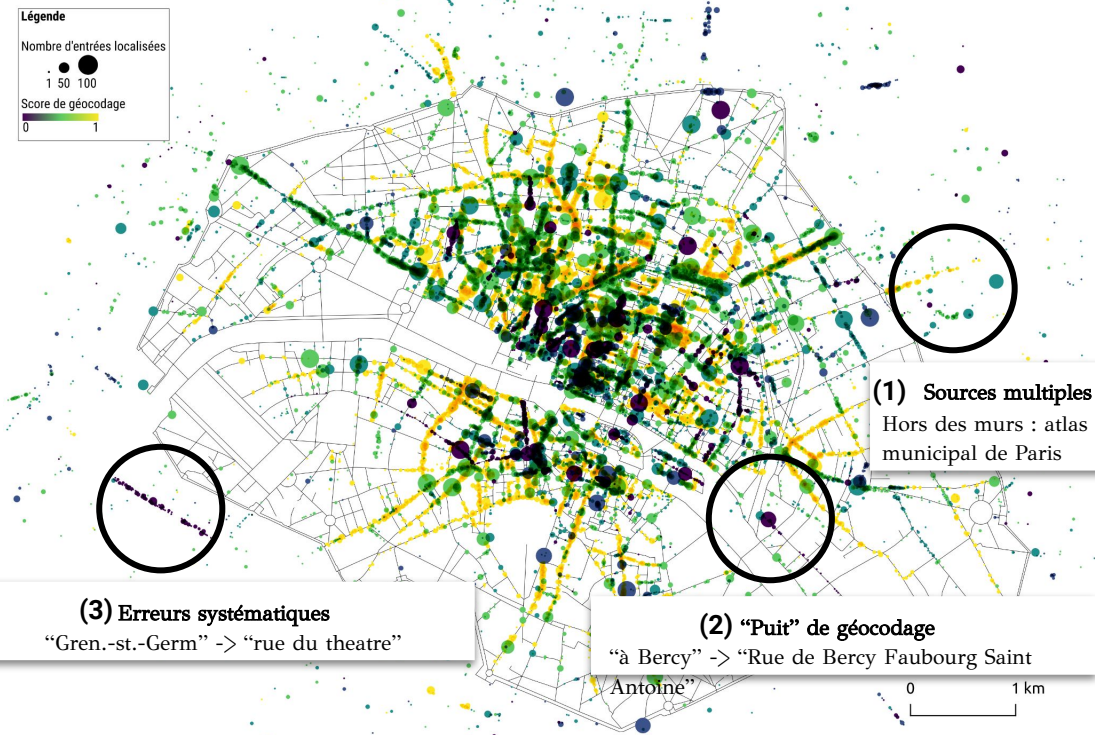
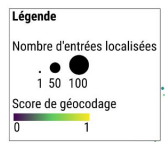
- des effets croisés: bruit d'extraction, complétude des données support, mécanismes de géocodage
- les annuaires "éloignés" temporellement des plans contiennent davantage d'adresses absentes des bases vectorisées

Adresses extraites par années de publications de annuaires et proportion géocodée.

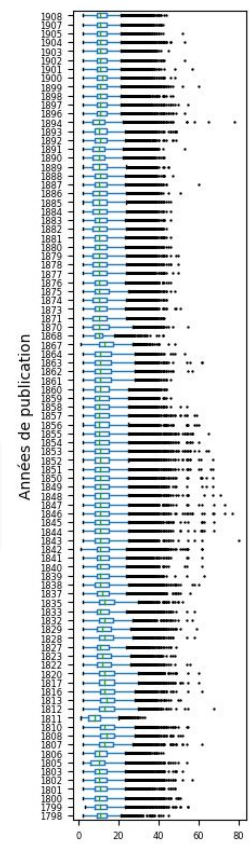


Résultats du géocodage des annuaires

Entrées géocodées pour l'année 1838 : Almanach du commerce de Paris[...] (Bottin)
et Annuaire général du commerce, de l'industrie et de l'agriculture de France [...] (Henrichs).



Longueurs des énoncés d'adresses en nombre de caractères et exemples typiques



Nombre de caractères

r. d. des, 123, che 38,
Bar 63, NAS, r. d, bis,
Tri, imp, BaC 104, Dou, St-, omb,
Bac 142., r.,
e N T, L Ver

Mauvaises-Paroles 15

quai aux Fleurs, au coin de la rue
de la Cité

avenue de Neuilly, sur la pelouse
près la rue du chemin de Versailles

Analyse critique

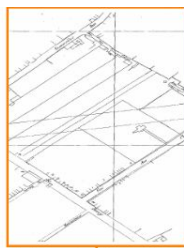
Vue globale des sources d'adresses, 1791-1937



Atlas de Verniquet



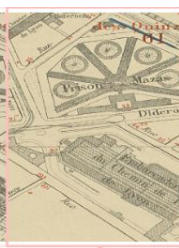
Cadastre



Atlas de Jacoubet



Atlas Mun. 1882



Atlas Mun. 1893



Atlas Mun. 1900



Atlas Mun. 1912



Atlas Mun. 1926



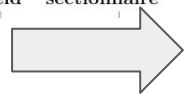
Atlas Mun. 1937



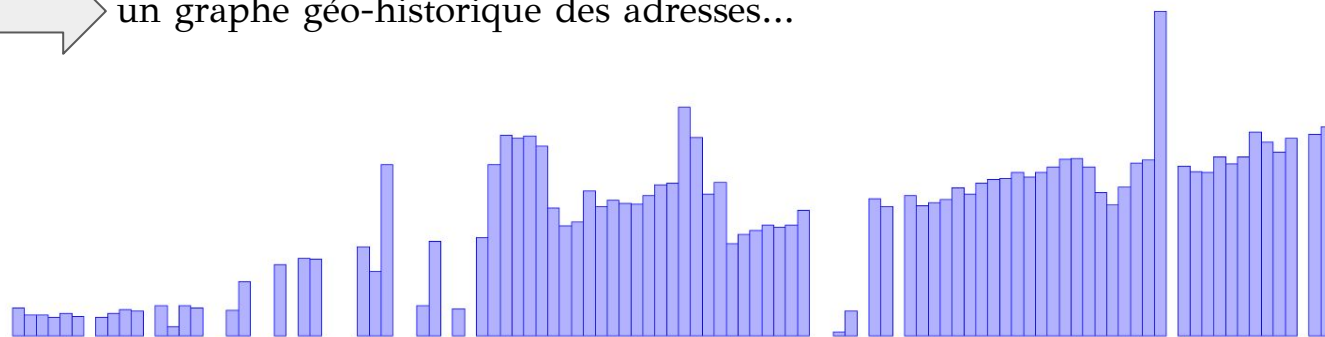
numérotations...
krenfeld sectionnaire

moderne

moderne révisée



un graphe géo-historique des adresses...



Adresses uniques

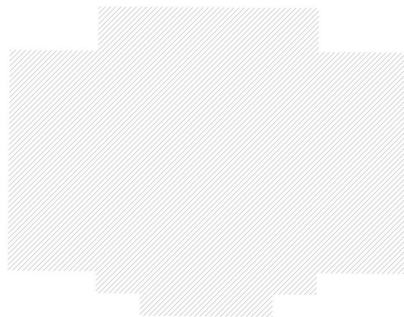
Les adresses constituent des mesures morphogénétiques

Emprises des levés

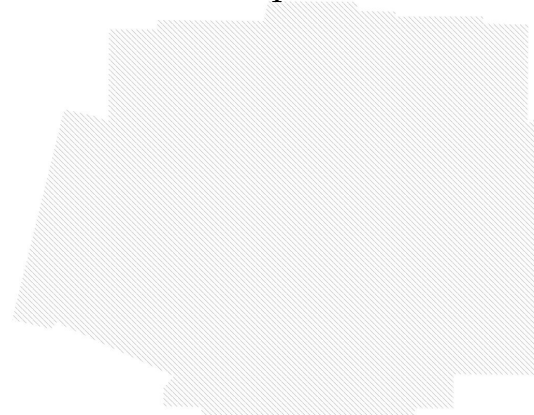
Atlas de Verniquet (~1791)



Atlas de Jacoubet (~1827)



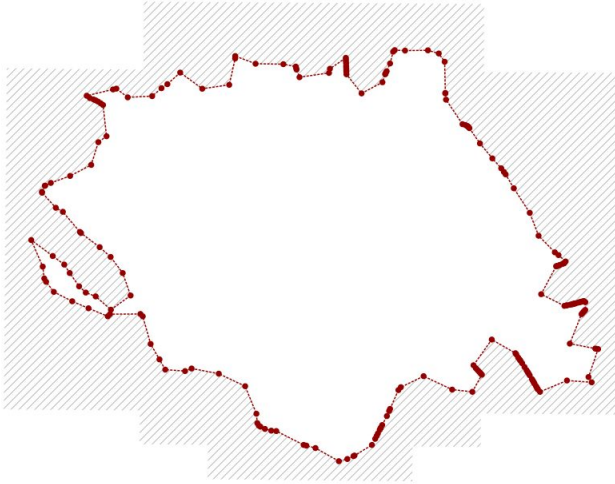
Atlas Municipal (~1888)



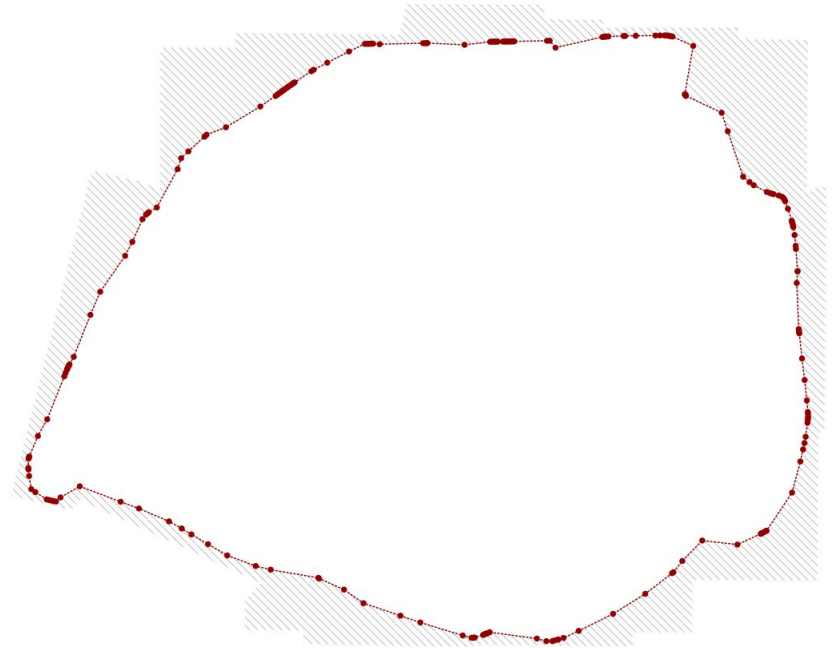
Les adresses constituent des mesures morphogénétiques

Fronts de numérotation

Atlas de Jacoubet (~1827)



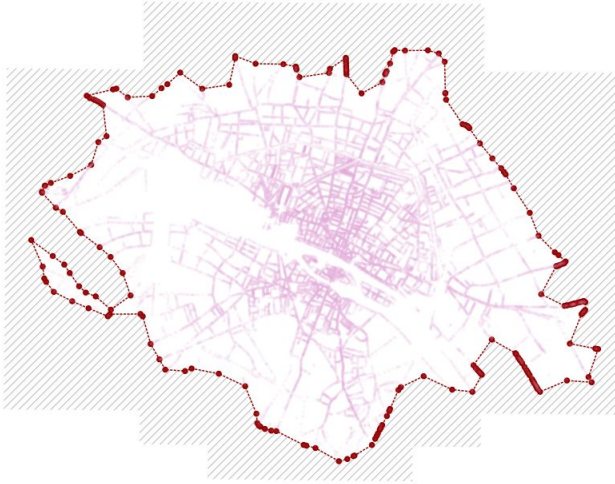
Atlas Municipal (~1888)



Les adresses constituent des mesures morphogénétiques

Ensemble des adresses relevées

Atlas de Jacoubet (~1827)



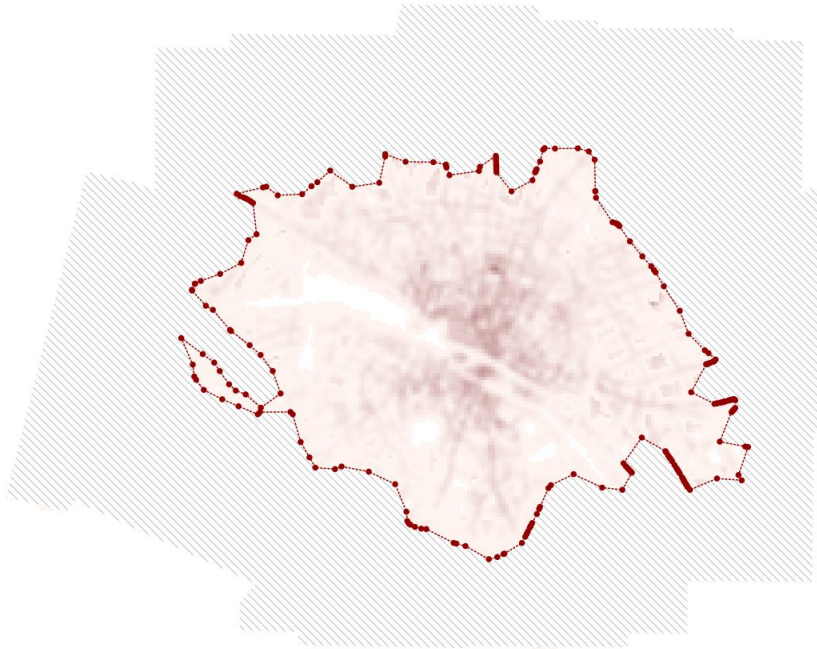
Atlas Municipal (~1888)



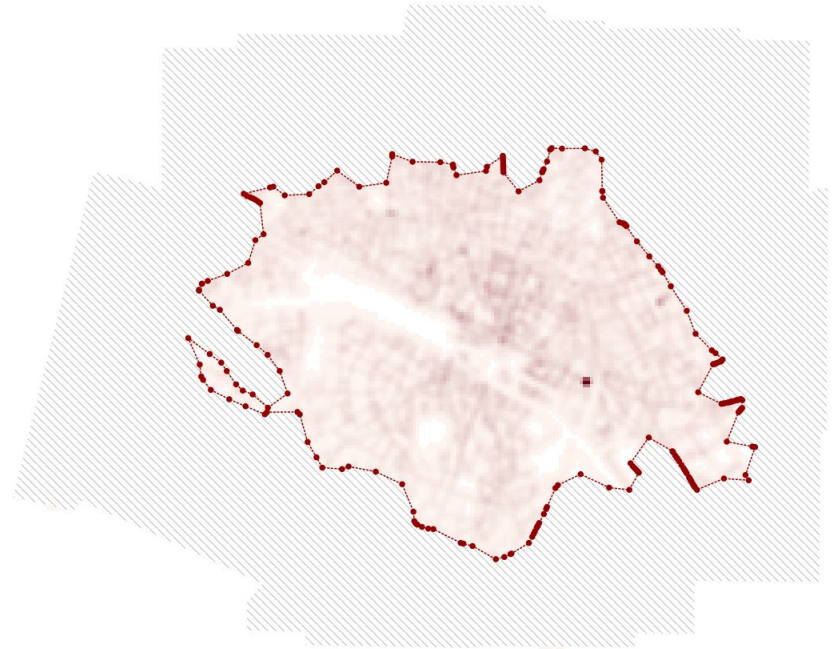
Les adresses constituent des mesures morphogénétiques

Densités d'adresses (carreaux de 50m)

Atlas de Jacoubet (~1827)

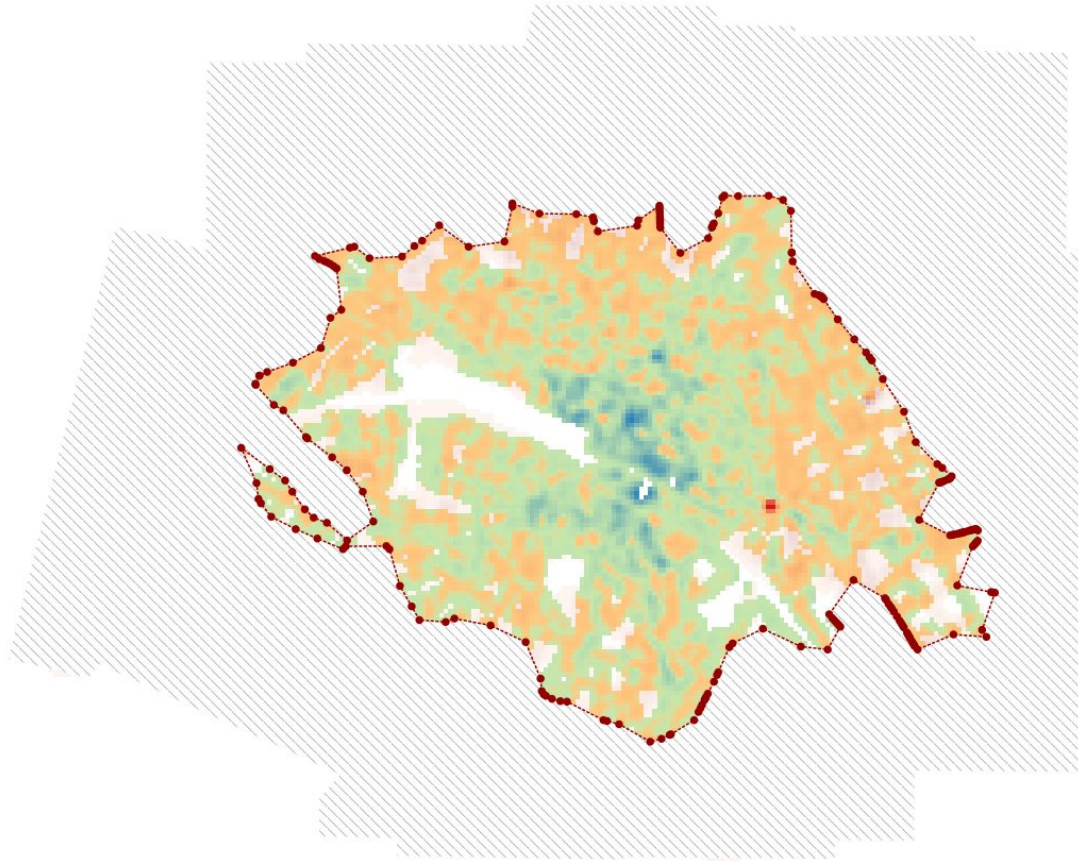


Atlas Municipal (~1888)



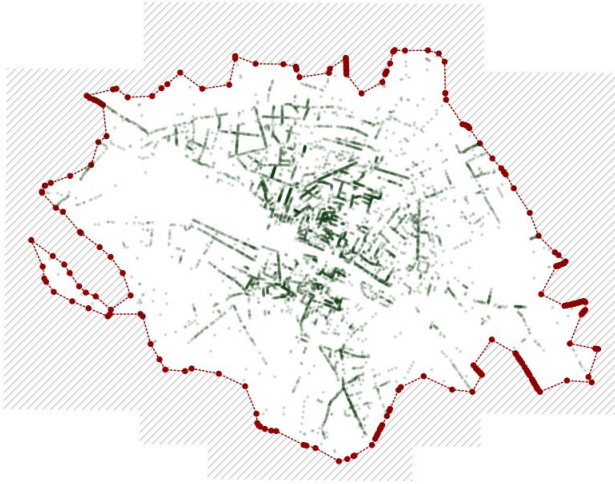
Les adresses constituent des mesures morphogénétiques

Différence des densités d'adresses (carreaux de 50m)

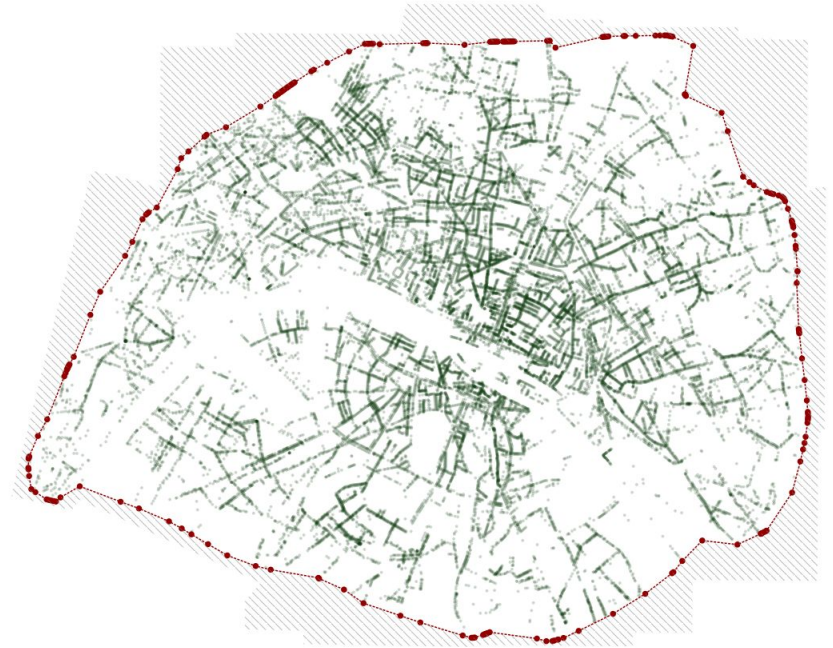


Les adresses utilisées dans les annuaires constituent des mesures morphogénétiques

Atlas de Jacoubet (~1827)



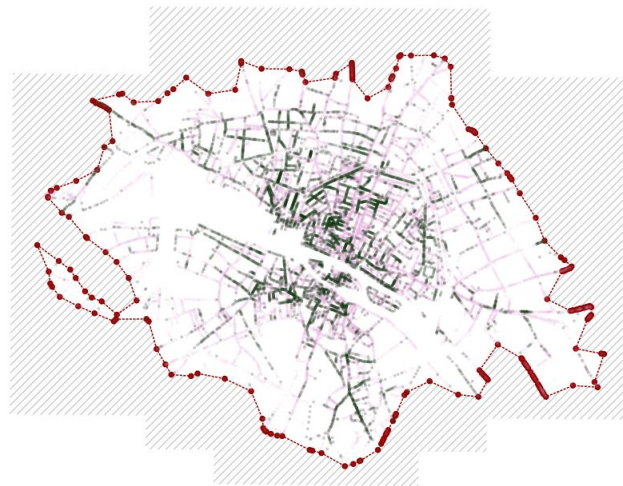
Atlas Municipal (~1888)



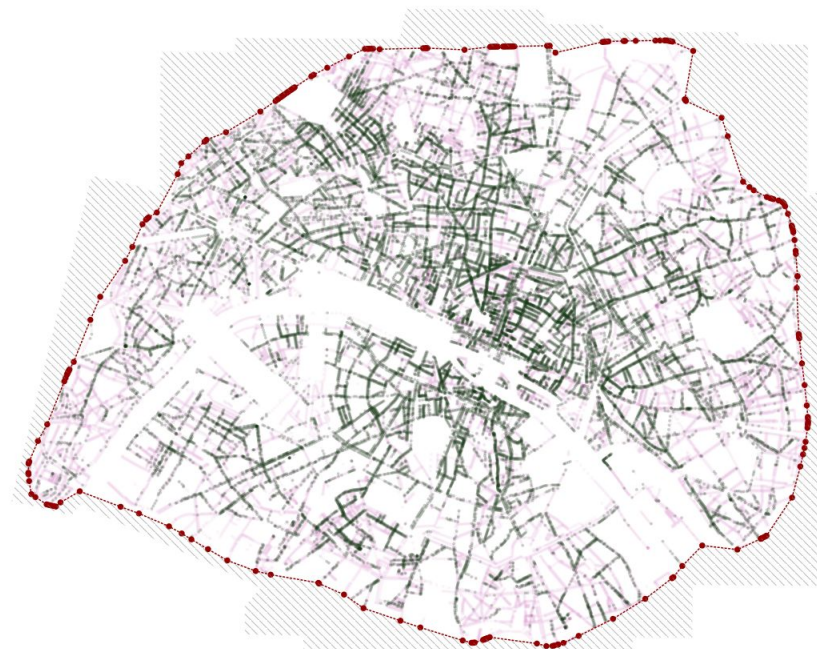
Les adresses utilisées dans les annuaires constituent des mesures morphogénétiques

Superposition avec les adresses relevées

Atlas de Jacoubet (~1827)



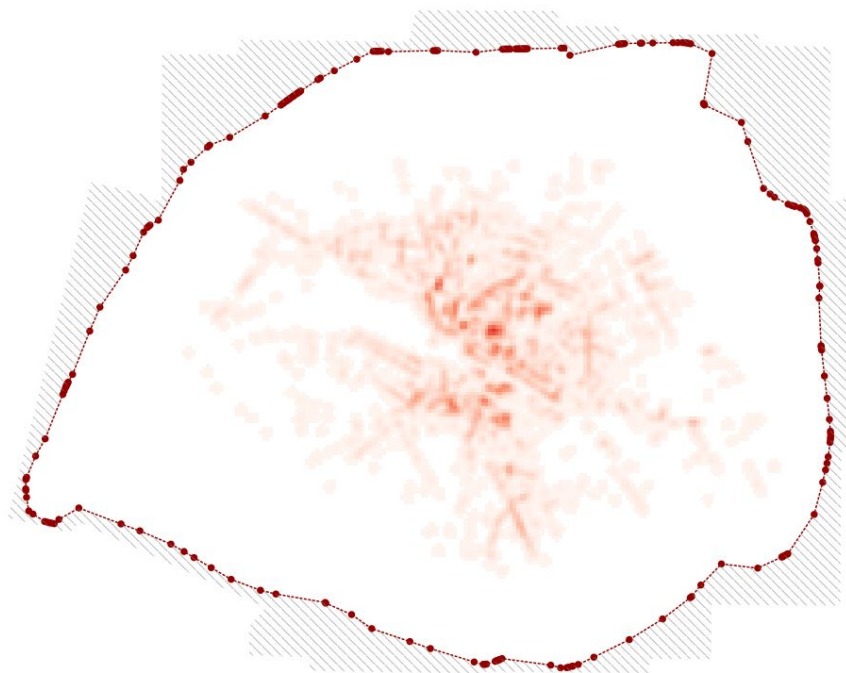
Atlas Municipal (~1888)



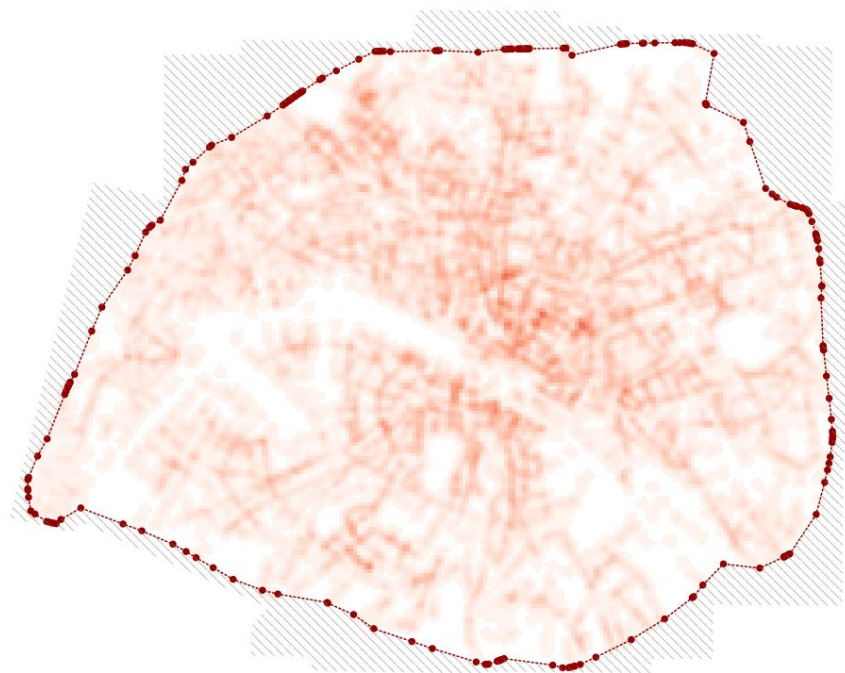
Les adresses utilisées dans les annuaires constituent des mesures morphogénétiques

Densités (carreaux de 50m)

Atlas de Jacoubet (~1827)

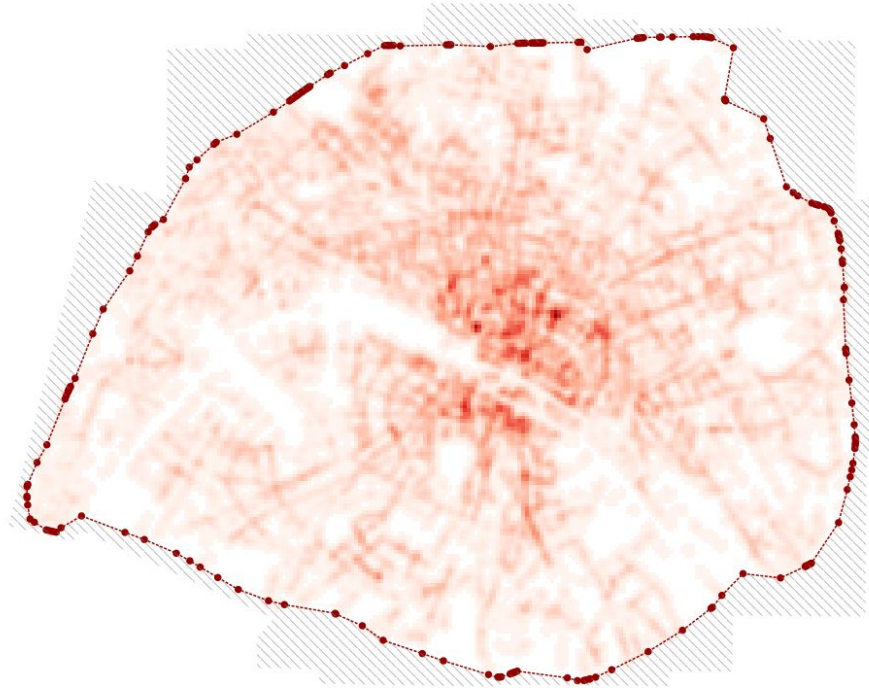


Atlas Municipal (~1888)



Les adresses utilisées dans les annuaires constituent des mesures morphogénétiques

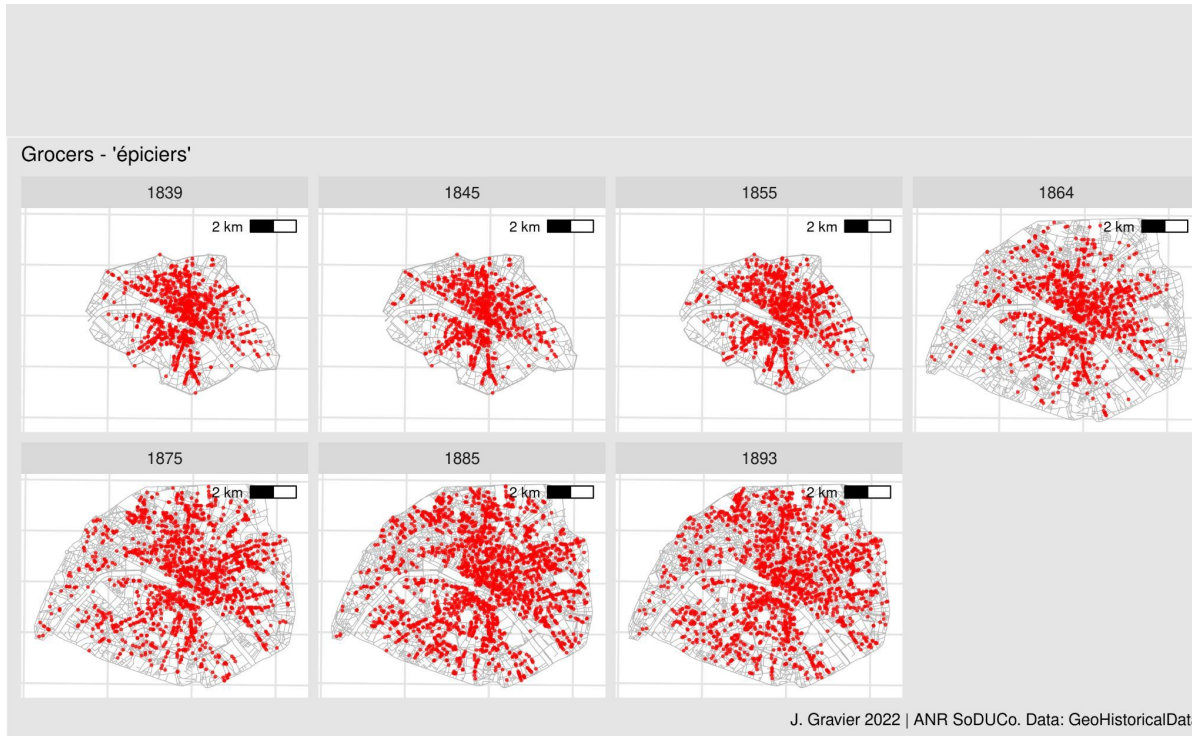
Densités (carreaux de 50m) : Atlas de Jacoubet (~1827) + Atlas Municipal (~1888)



Des premiers éléments de l'évolution des activités parisiennes au fil du siècle

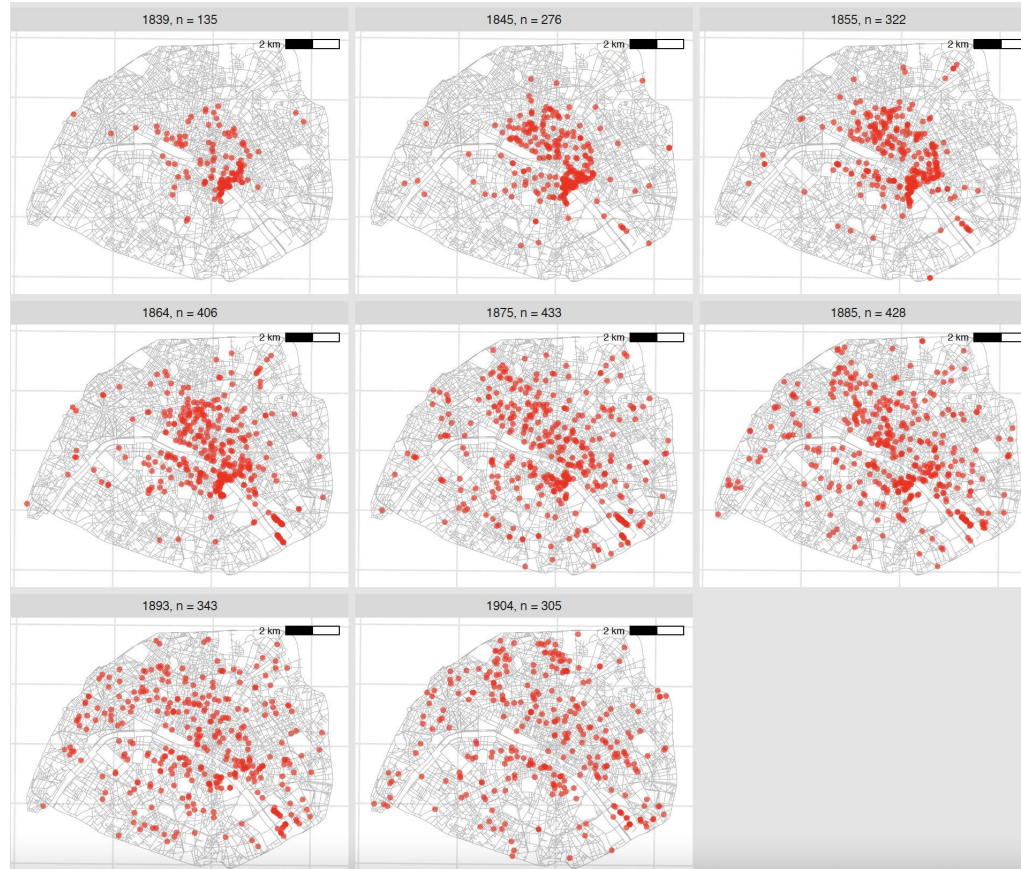


Des premiers éléments de l'évolution des activités parisiennes au fil du siècle



Des premiers éléments de l'évolution des activités parisiennes au fil du siècle

les marchands de vins en gros



Conclusion

Approche scientifique du projet SoDUCo

Analyse critique croisée de 2 collections

Atlas et annuaires

Exploitation des **redondances spatiales et temporelles**

Détection d'erreurs ou de changements, stabilisation de la reconnaissance

Extraction d'information **pragmatique**

Assister l'utilisateur expert, viser l'annotation collective dans une certaine mesure

Compromis entre **généricité** et **spécificité**

Viser une solution fonctionnelle sur notre corpus, avec des composants réutilisables

Free/open, accessible, indexé, reproductible

[Dépôts GitHub](#), <https://geohistoricaldata.org>, jeux de données, publications...

Perspectives

Outils libres d'extraction d'information

dans les **cartes** (reconstitution et géoréférencement d'atlas)

dans les **annuaires**

Données ouvertes

évolution morphologique de Paris

évolution des activités professionnelles de Paris

un **graphe géo-historique des adresses** de Paris

Catalogue pour la diffusion d'un *corpus cohérent*

interopérabilité (complémentaire de gallica, data.bnf, archives, etc.) - *formats ? outils ?*

traçabilité (données et traitements) - *formats ? outils ? bonnes pratiques ?*

Approche critique et analyse de la **co-évolution**

Un cycle de séminaires SoDUCo-BnF

https://soduco.github.io/soduco_bnf_seminars/

Séance 1/4 (11/04/2022, BnF Tolbiac)

Paris, capitale géodésique. Atlas et plans à grande échelle: 1780 - 1950

Séance 2/4 (10/11/2022, BnF Tolbiac)

Res(t)ituer les adresses des almanachs et annuaires commerciaux parisiens du XIXe siècle.

Séance 3/4 (23/05/2023, BnF Richelieu)

Analyses spatiales formalisées de données géo-historiques.

Matinée : Les chaînes de traitement des documents cartographiques des 18e et 19e siècles.

Après-midi : Analyses spatiales formalisées de données géo-historiques

Séance 4/4 (octobre/novembre 2023)

Données ouvertes, outils libres, science collaborative et participative

Resources

Julie Gravier, Marc Barthelemy. Exploitation des données Soduco : premiers éléments de l'évolution des activités parisiennes au fil du siècle. *2e Journée d'étude SoDUCo/BnF. Res(t)ituer les adresses des almanachs et annuaires commerciaux parisiens du XIXe siècle. Un corpus de localisations urbaines à grande échelle*, BnF Bibliothèque nationale de France, Nov 2022, Paris, France. [⟨hal-03961814⟩](#) [Code](#) [Data](#)

Nathalie Abadie, Edwin Carlinet, Joseph Chazalon, and Bertrand Dumenieu. *A Benchmark of Named Entity Recognition Approaches in Historical Documents Application to 19 Century French Directories*. In: Uchida, S., Barney, E., Eglin, V. (eds) Document Analysis Systems. DAS 2022. Lecture Notes in Computer Science, vol 13237. Springer, Cham. Proceedings of the Document Analysis Systems international workshop, held in La Rochelle, France, 22-25 may 2022. [⟨10.1007/978-3-031-06555-2_30⟩](#). [Code](#) [Data](#)

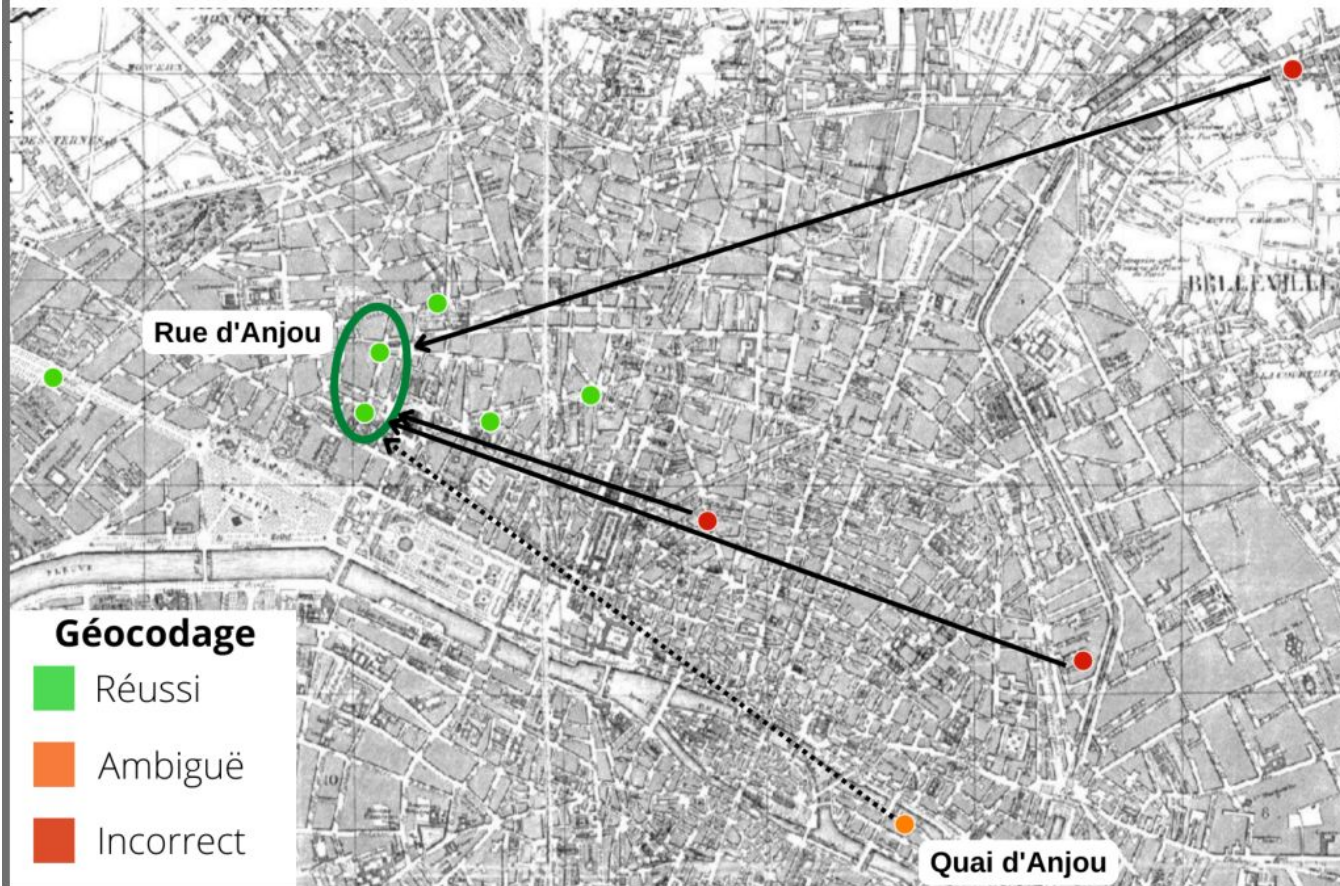
Yizi Chen, Edwin Carlinet, Joseph Chazalon, Clément Mallet, Bertrand Duménieu, et al.. Combining Deep Learning and Mathematical Morphology for Historical Map Segmentation. *Discrete Geometry and Mathematical Morphology, First International Joint Conference, DGMM 2021, Uppsala, Sweden, May 24–27, 2021, Proceedings*, 12708, pp 79 - 92, 2021, Lecture notes in computer science, [⟨10.1007/978-3-030-76657-3_5⟩](#). [⟨hal-03101578⟩](#) [Code](#) [Data](#)

Yizi Chen, Edwin Carlinet, Joseph Chazalon, Clément Mallet, Bertrand Dumenieu, et al.. Vectorization of historical maps using deep edge filtering and closed shape extraction. *16th International Conference on Document Analysis and Recognition (ICDAR'21)*, Sep 2021, Lausanne, Switzerland. pp.510-525, [⟨10.1007/978-3-030-86337-1_34⟩](#). [⟨hal-03256073⟩](#) [Code](#) [Data](#)

Bertrand Dumenieu, Julien Chadeyron, Pascal Cristofoli, Julien Perret, Laurence Jolivet, et al.. Engraved footprints from the past. Retrieving cartographic geohistorical data from the Cassini Carte de France, 1750-1789. *International Cartographic Conference 2019*, Jul 2019, Tokyo, Japan. [⟨hal-02173847⟩](#) [Data](#)

Benoit Costes, Julien Perret. A hidden Markov model for matching spatial networks. *Journal of Spatial Information Science*, 2019, 18, [⟨10.5311/JOSIS.2019.18.489⟩](#). [⟨hal-03545163⟩](#) [Code](#) [Data](#)

Visualisation cartographique



Ateliers de la famille
Tournachon - Nadar
entre 1850 et 1908

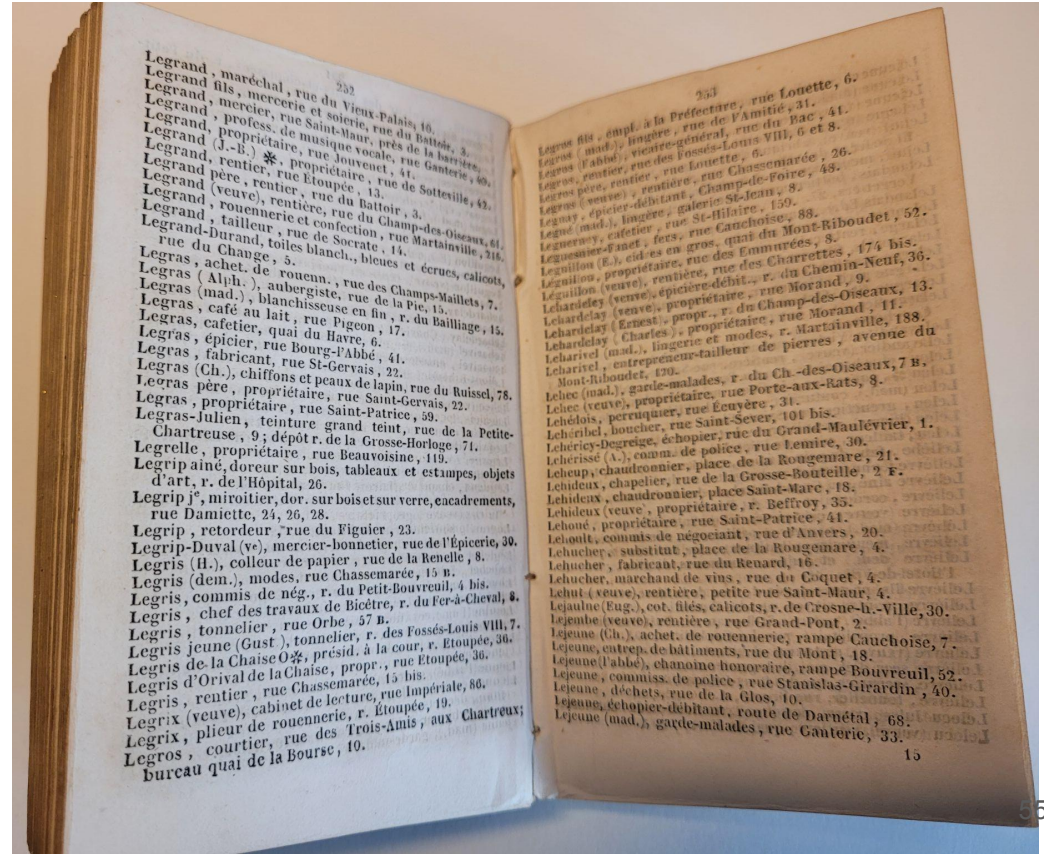
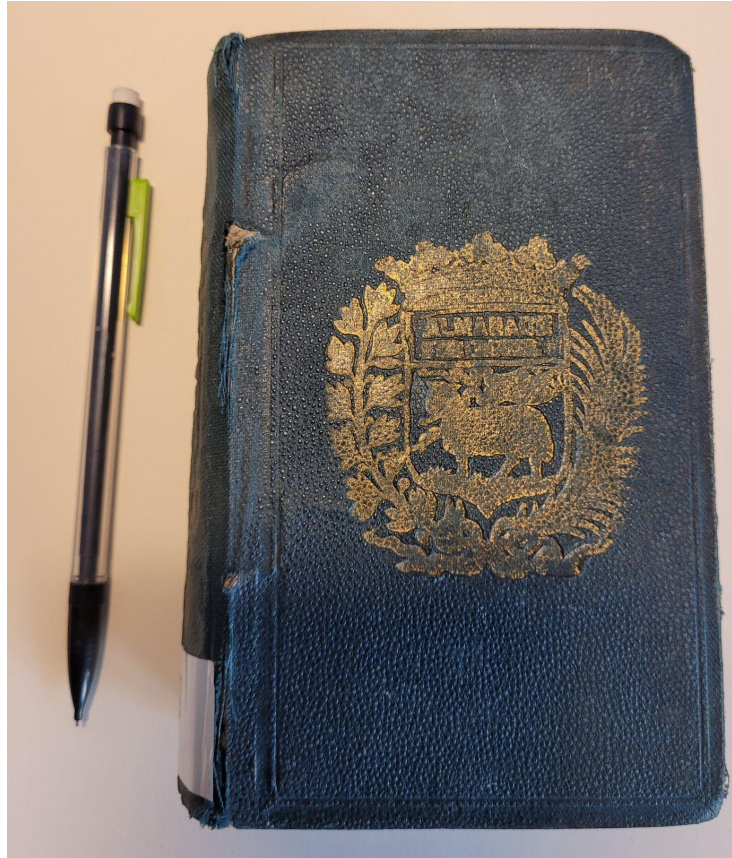


https://solenn-tl.github.io/stage_demo_photographies/

Partie 2 : Focus sur l'analyse des annuaires

*Travaux menés par Nathalie **Abadie** (IGN), Edwin **Carlinet** (EPITA), Joseph **Chazalon** (EPITA), Bertrand **Duménieu** (EHESS) et Solenn **Tual** (IGN),
avec l'aide de Stéphane **Baciocchi** (EHESS), Pascal **Cristofoli** (EHESS) et Julien **Perret** (IGN).*

Annuaire du commerce de Paris, 1798-1861



Annuaire du commerce de Paris, 1798-1861

Un corpus de **141** annuaires numérisés.

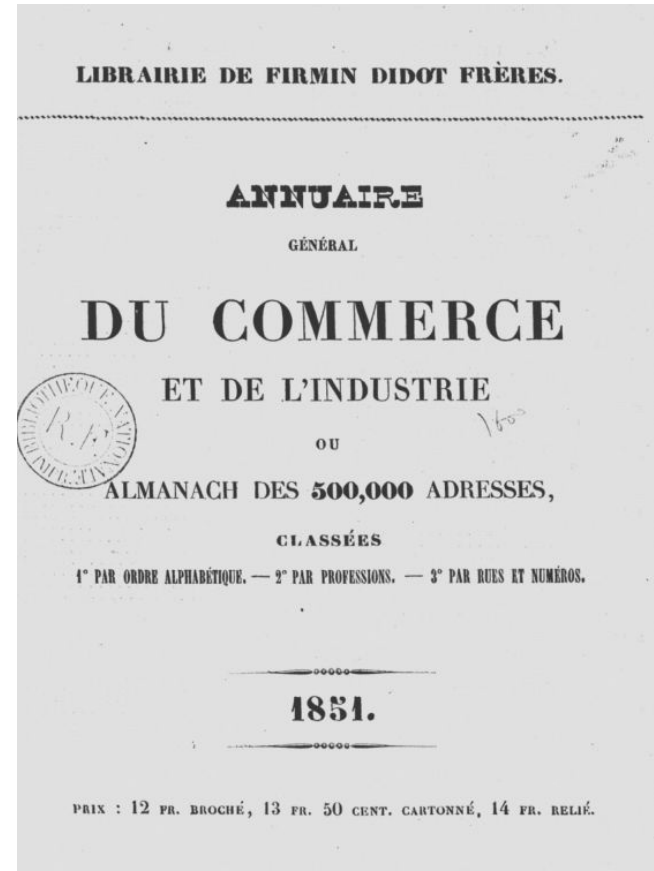
- **Centaines de milliers de pages**
- **Millions d'entrées brutes**

1 entrée = (patronyme/raison soc., activité, adresse, autres info.)

Focus sur les **3 sections principales** :

1. Index par « patronyme » (raison sociale)
2. Index par « professions » (activité socio-éco.)
3. Index par rue et numéro

Nombreuses autres sections et informations.



lons ronds et demi-ronds et carvés, et de toute sorte de moulure ; spécialité de fabrication de broches en acier et cuivre pour les peignes à tisser draps, voiles et couvertures, toile métallique pour tout ce qui concerne le tissage, M. H. 1849, Amandiers-Popincourt, 19.

Auger (Vve), mécanicien, fabr. d'emporte-pièces pour festons de garnitures de robes et de mantelets, emporte-pièces pour cartonnage, semelles de souliers, etc., tient un assortiment de maillets, billots et plombs, pass. de la Trinité, 77.

Avoyné-Bainée, serrurier-mécan., fab. de lits en f. r. 1839-44, 1819, Boulangers-St-Victor, 22.

Bijard, quai Jemmapes, 248.

Baptrosse père, tout ce qui a rapport à la fabrique d'indiennes, Charenton, 58.

Barbot, Popincourt, 58.

Bardiés, Trois-Bornes, 21.

Bardou, Descartes, 38.

Bariquand, horloger-mécanicien, entreprend toute espèce de fabrication nouvelle, quelle que soit la précision demandée, et construit toutes sortes de machines et outils pour Messieurs les inventeurs; vis cylindriques et coniques, découpage à façon, fendage de roues; pièces détachées fabriquées par procédés mécaniques pour toute espèce d'industrie; 1849, Ecouffes, 20. au Mar., près le marché des Blancs-Manteaux. *

Baudat, constructeur de mécaniques à scier le placage, le bois en grume, circulaires à cylindre, à scier l'ivoire, le marbre et la pierre; les machines sont mues par manège, eau ou vapeur; garantit la bonne exécution de ces machines, se charge de la pose et mise en activité, fille toutes sortes de vis et d'écrous, tourne au chariot et autrement, fend les engrenages à la plate-forme, dentelle les scies à la machine, fait des envois en province et à l'é-

Becker, mécanicien-graveur, breveté d'invention (sans garantie du gouvernement), fabrique spéciale de presses à copier, presses à timbre sec et bumide, presses à cacheter, presses à percussion pour l'extraction des matières végétales et pour le satinage des papiers, composteurs pour papiers à lettres, pour raisons de commerce, griffes de toutes sortes, et généralement toute la gravure sur métaux, St-Denis, 380, passage Lemoine. *

Beerstecher (A.), Loureine, 88

Berendorff (J.), M. H. 1844, Mouffetard, 294, atelier de construction de machines à vapeur et autres, S. E., 1844, 1849.

Bergeron, 1849, Marais-du-Temple, 73.

Bernier aîné, outils établis, presses, 1844, Faubourg-St-Antoine, 89.

Bernier, Ménéilmontant, 90.

Berthet, Simon-le-Franc, 15.

Berton, Neuve-St-Denis, 12.

Bertrand, mécanicien; breveté (sans garantie du gouvernement), construit toutes espèces de machines et outils, spécialité de machines pour les fabricants de matières premières de chapellerie, machines à vapeur et tours pour les fabricants de chapeaux; tondeuses et souffleuses de poils de lièvres et de lapins, réparation et entretien de machines à vapeur; on trouve toujours chez lui des machines à vapeur toutes prêtes à être mises en place, Vieille-du-Temple, 58, anc. 72. *

Beslay (Ch.), 1839, 1849, Neuve-Popincourt, 17.

Bettinger, mécanicien, tient fabrique générale de toutes sortes de serrures de nécessaires, de portefeilles, ainsi que loqueteaux et charnières en cuivre, vis, et fait tout ce qui concerne son état, Vieilles-Etuves-St-Martin, 9, ci-devant St-Martin, 104, passage de la Réunion. 3. *

Bianchi (By.), 1844 et 1840

MÉCANICIENS.

province, à des prix très-modérés, St-Jacques, 261. *

Boscho aîné, ingénieur-mécanicien, inventeur de divers perfectionnements au métier Jacquard, S. E. 1850, Amandiers-Popincourt, 22, ci-devant St-Maur, 14. *

Boscher, Bondy, 76.

Bosquillon, 1823-27-34, constructeur de perçages accélérés et de mécaniques Jacquard parisiennes, Paradis-Poissonnière, 20.

Boucher, Saint-Pierre-Popincourt, 18.

Bouchon, dépôt de moulins à bras portatifs, Nve-St-Nicolas, 16.

Bouhey (Ene), mécanicien, fab. de machines à couper les peaux de lapins, souffleuses de toutes dimensions marchant à bras et à la vapeur, tours de chapeliers, 1849, Francs-Bourgeois-Marais, 3, ci-devant Beaubourg, 59. *

Bouhon, appareil dit cale à fléau, breveté sans garantie du gouvernement, destiné au soulagement des chevaux dans les montées, 1849, Société d'encouragement 1850, pl. Dauphine, 71. *

Bourdier, bandagiste-hermier, mécanicien-orthopédiste, fournisseur de la Société protestante de prévoyance et de secours mutuels de Paris, fabrique spéciale de toutes espèces de ressorts et de bandages, béquilles, appareils orthopédiques, ceintures hypogastriques, bas lacés en tous genres, suspensoirs, objets d'allaitement, et tous articles en gomme élastique, Cadran, 5.

Bourdon (Eug.), construction de machines à vapeur à haute, moyenne ou basse pression, pompes, machines à élever l'eau, engrenages, arbres de couche, machines à vapeur locomobiles pour les irrigations et les exploitations rurales; breveté (sans gar. du g.) pour de nouveaux manomètres métalliques autorisés par décision ministérielle des travaux publics, en date du 6 novembre 1849, pour

gner, mécanique nouvelle, presse à copier les lettres, perfectionnée 1849, Saint-Jacques, 169, près le Panthéon. *

Brocard (Léon), machines à papier, St-Gilles, 12.

Burckel, fab. d'écrites, fuseaux et broches, Nve-de-Lappo, 3.

Busser, fab. de presses lithographiques, machines à réparer, fait les réparations à des prix modérés, Lamartine, 18-20, ci-devant Coquenard. *

Butt, Buisson-St-Louis, 16.

Cabanis et Cie, association pour la construction des machines à élever les eaux, pompes domestiques, moteurs hydrauliques de toute nature, machines soufflantes pour les forges, machines à eaux de Seltz, presses à découper, outillage et réparation de machines, Vainigriers, 32.

Cahouet, breveté (sans garantie du gouvern.) pour les porte-moules à bougies emboutis d'un seul morceau, et les moules montés sans aucune soudure, et pouvant les monter et les démonter soi-même; fabricant de mandrins, fondeur de moules de chandelles, bougies, cierges, spécialité, 1849, Marché-aux-Veaux, 4.

Calamel, Dupetit-Thouars, 16.

Calard (T.-F.), fab. de feuilles métalliques percées en fer, cuivre et zinc; tôle en râpe pour le nettoyage des blés, cribles à main pour remplacer la peau.—Atelier de construction de machines à percer, à cintrer les fers à froid et à ajuster les ressorts de voitures, Leclerc, 8, Faub.-St-Jacques, 1849.

Calla fils, ingénieur-mécanicien-fond.; machines pour travailler les métaux, moteurs hydrauliques, moulins à blé, 1839, 1844, 1849, Lafayette, 11, et Chabrol, 20, à La-Chapelle.

Cambray, membre de la Société d'Encouragement, de la Société nationale et centrale d'Agriculture, et de la Société de l'Académie de l'Industrie, fabrique

Un corpus hétérogène

Ouvrages produits par des éditeurs variés, numérisés par différents prestataires.

Des différences de

- Longueur, richesse et structuration de l'information
- Mises en page, fontes
- Qualité inégale : papier, résolution des images, présence de marques...

Non-Commerçans. (Paris). 269

<p>Chardin, R. Pavée, 16. — R. G. Chardin, R. Michel Lepelletier, 21. Chardon, (Ve.) R. S. Marc, 15.</p>	<p>Cheviron, R. Chapou, 13. Chimay, (Mme.) R. de Varennes, 31. Choart-Duplessis, R. de Turanne, 31.</p>
---	--

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

DARRAS (Thomas), r. de la Vieille-Monnaie, 10. GALLIENNE j^e, r. de la Heumerie, 3.
 Briquets et veilleuses, mèches à quinquets, à quinquet, veilleuses mèches, souffrées ; mèches souffrées, pierres, agaric de chêne, pierres, agaric, bouchons, liège.
 Liège en planches, bouchons. LEROY, r. Aubry-le-Boucher, 43.

<p>BAUDOYER (place). IX Arr. Hôtel-de-Ville. ← Rue Tixeranderie, pourtour St-Gervais, Saint-Antoine et Renaud-LeFèvre.</p> <p>1 Lissoty (Vve), vins. 2^e Privé, distillateur. Lemoine-Cluzel et Leroy, nouveautés. Chantrier, court.-gourm.</p>	<p>26^e Longpré aîné, bijoutier en or et argent. Saint-Omer, émailleur. Cellier (A.), graveur-ci- scieur.* Rousseau (J.), bijoutier en or.* Benoit, orfèvre-fabr. Léréty, doreur. 30 Bouton, fab. de cuir ver- nis.* 31 Pardon, vins.</p>	<p>Bourguille, fabr. de presses. Vautain, passementier. Fininojac, bronze doré. Rabé aîné, fabr. de bal- lons.* Gaulin, chapelier. Moisy, tabletier. 49^e Cendrier aîné, prop. Desmarests, fab. bottes d'emballage. Ferrand, lapidaire.</p>	<p>7 Ecole communale de jeu- nes filles. Berthelot, vins. 6 Verstaen, serrurier-mé- canicien. 8 Michel, brossier. 9 Labottiere, serrurier. 10 Sacrez, vins. 12 Baudoin, épici. 13 Lejard, clouteries et crê- pins. 14 Badauel (Vve), fab. de</p>	<p>et tapisseries. 10 Lainé jeune, vins. Jumelles omnibus et en- treprise générale des Omnibus. 11 Melouzat, vins en gros, et à Bercy, Port, 31. 12 Combaud, coiffeur. Monmain (P.), vins en gros. 13 Dufailly, sculpt. fabr. de carton-pierre.</p>
---	--	--	---	--



Un instantané des acteurs commerciaux à un instant

Bibliothèque Ste-Geneviève, Clotilde, 1.
Bibliothèque de la Ville, quai d'Austerlitz, 33 (provisoirement).
Bibl'que 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.

Didot 1841a - page 95

Bibliothèque Ste-Geneviève, Clotilde, 1.
Bibliothèque de la Ville, quai d'Austerlitz, 33 (provisoirement).
Bibl'que 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.
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.

Didot 1842a - page 117

Bibliothèque Ste-Geneviève, 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 1843a - page 129

Bibliothèque Ste-Geneviève, 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.
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 1844a - pages 125-126

Un instantané des acteurs commerciaux à un instant

Grande **redondance** d'information intra- et inter-annuaires.

Bibliothèque Ste-Geneviève, Clotilde, 1.
Bibliothèque de la Ville, quai d'Austerlitz, 33
 (provisoirement).
Bibl'que 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.
 Bican (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-Mar-
 tin, 45.

Didot 1841a - page 95

Bibliothèque Ste-Geneviève, Clotilde, 1
Bibliothèque de la Ville, quai d'Austerlitz, 33
 (provisoirement).
Bibl'que 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.
 Bican (Vve) et fils, fondeur en cuivre, place
 de la Corderie-du-Temple, 26.
 Bichel, épicier, marché d'Aguesseau, 15.
 Bichard (Mme), Nve-de-Luxembourg, 17.
 Bichard, tabacs et eau-de-vie, Faub.-St-Mar-
 tin, 45.

Didot 1842a - page 117

*Bibliothèque Ste-Geneviève, 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.
 Bichel, épicier, Marché-d'Aguesseau, 15.
 Bichard, tabac et eau-de-vie, Faub.-St-Mar-
 tin, 45.

Didot 1843a - page 129

*Bibliothèque Ste-Geneviève, 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-Plan-
 tes, Carier, 20.
 Bibus, tailleur, Richelieu, 31.
 Bical, fab. de jouets, Montmorency, 33.
 Bican (Vve) et fils, fondeurs en cuivre, place
 de la Corderie-du-Temple, 26.
 Bichel, épicier, Marché-d'Aguesseau, 15.
 Bichard, tabac et eau-de-vie, Faub.-St-Mar-
 tin, 45.

Didot 1844a - pages 125-126

Un instantané des acteurs commerciaux à un instant

Des entrées **disparaissent** au cours du temps.

Bibliothèque Ste-Geneviève, Clotilde, 1.
Bibliothèque de la Ville, quai d'Austerlitz, 33
 (provisoirement).
Bibl'que 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.
 Bican (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-Mar-
 tin, 45.

Didot 1841a - page 95

Bibliothèque Ste-Geneviève, Clotilde, 1.
Bibliothèque de la Ville, quai d'Austerlitz, 33
 (provisoirement).
Bibl'que 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.
 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-Mar-
 tin, 45.

Didot 1842a - page 117

Bibliothèque Ste-Geneviève, 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-Mar-
 tin, 45.

Didot 1843a - page 129

Bibliothèque Ste-Geneviève, 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-Plan-
 tes, Cuvier, 29.
 Bibus, tailleur, Richelieu, 31.
 Bical, fab. de jouets, Montmorency, 33.
 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-Mar-
 tin, 45.

Didot 1844a - pages 125-126

Un instantané des acteurs commerciaux à un instant

D'autres apparaissent.

Bibliothèque Ste-Geneviève, Clotilde, 1.
Bibliothèque de la Ville, quai d'Austerlitz, 33
 (provisoirement).
Bibl'que 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.
 Bican (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-Mar-
 tin, 45.

Didot 1841a - page 95

Bibliothèque Ste-Geneviève, Clotilde, 1.
Bibliothèque de la Ville, quai d'Austerlitz, 33
 (provisoirement).
Bibl'que 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.
 Bican (Vve) et fils, fondeur en cuivre, place
 de la Corderie-du-Temple, 26.
 Bichel, épicier, marché d'Agnesseau, 15.
 Bichard (Mme), Nve-de-Luxembourg, 17.
 Bichard, tabacs et eau-de-vie, Faub.-St-Mar-
 tin, 45.

Didot 1842a - page 117

Bibliothèque Ste-Geneviève, 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.
 Bichel, épicier, Marché-d'Agnesseau, 15.
 Bichard, tabac et eau-de-vie, Faub.-St-Mar-
 tin, 45.

Didot 1843a - page 129

Bibliothèque Ste-Geneviève, 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-Plan-
 tes, Cuvier, 29.
 Bibus, tailleur, Richelieu, 31.
 Bical, fab. de jouets, Montmorency, 33.
 Bican (Vve) et fils, fondeurs en cuivre, place
 de la Corderie du Temple, 26.
 Bichel, épicier, Marché-d'Agnesseau, 15.
 Bichard, tabac et eau-de-vie, Faub.-St-Mar-
 tin, 45.

Didot 1844a - pages 125-126

Un instantané des acteurs commerciaux à un instant

Et certaines changent.

Bibliothèque Ste-Geneviève, Clotilde, 1.
Bibliothèque de la Ville, quai d'Austerlitz, 33
 (provisoirement).
Bibl'que 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.
 Bican (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.

Didot 1841a - page 95

Bibliothèque Ste-Geneviève, Clotilde, 1.
Bibliothèque de la Ville, quai d'Austerlitz, 33
 (provisoirement).
Bibl'que 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.
 Bican (Vve) et fils, fondeur en cuivre, place
 de la Corderie-du-Temple, 26.
 Bichel, épicier, marché d'Aguesseau, 15.
 Bichard (Mme), Nve-de-Luxembourg, 17.
 Bichard, tabacs et eau-de-vie, Faub.-St-Martin, 45.

Didot 1842a - page 117

*Bibliothèque Ste-Geneviève, 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.
 Bichel, épicier, Marche-d'Aguesseau, 15.
 Bichard, tabac et eau-de-vie, Faub.-St-Martin, 45.

Didot 1843a - page 129

*Bibliothèque Ste-Geneviève, 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.
 Bichel, fab. de jouets, Montmorency, 33.
 Bican (Vve) et fils, fondeurs en cuivre, place
 de la Corderie-du-Temple, 26.
 Bichel, épicier, Marché-d'Aguesseau, 15.
 Bichard, tabac et eau-de-vie, Faub.-St-Martin, 45.

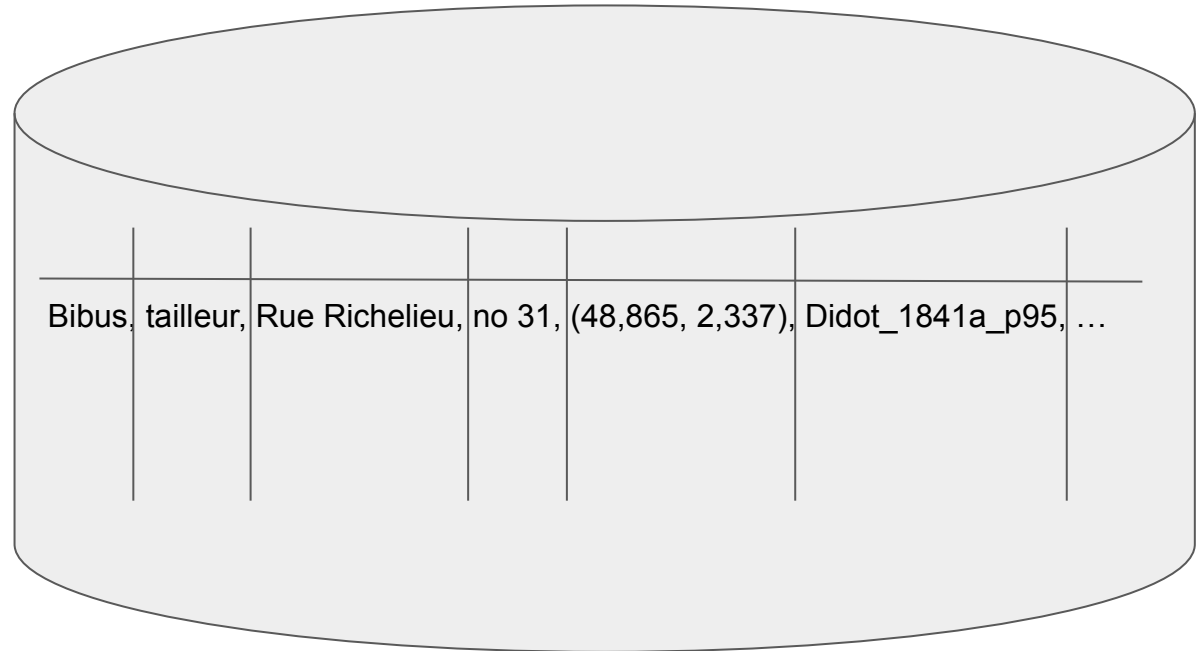
Didot 1844a - pages 125-126

Notre objectif d'extraction de données

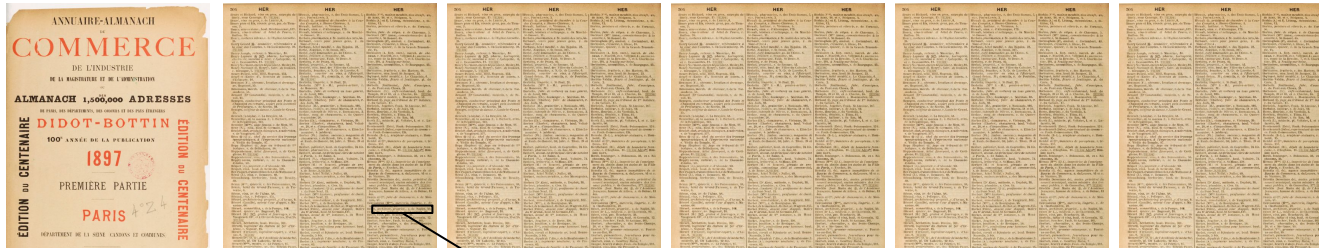
Extraire les entrées des annuaires pour les dédoublonner, les corriger, les relier et les géo-référencer.

Bibliothèque Ste-Geneviève, Clotilde, 1.
Bibliothèque de la Ville, quai d'Austerlitz, 33
(provisoirement).
Bibl'que 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.
Richard (Mme), Nve-de-Luxembourg, 17.
Richard, tabacs et eau-de-vie, Faub.-St-Martin, 45.

Didot 1841a - page 95



Des documents numérisés aux données structurées



Pages d'annuaires numérisées

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

Identification des entrées

Entrée d'annuaire

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

Transcription du texte

Texte numérique de l'entrée

PERSONNE ACTIVITÉ RUE NUMÉRO RUE NUMÉRO
Lemonnyer, plomberie, r. de Bondy, 86, et r. Bouchardon, 1.

Annotation du texte

Entités nommées

Structuration des entrées



Base de données intégrée

Besoin de reconnaissance d'entités nommées

Permet l'extraction automatique des contenus pertinents dans chaque entrée

Dartigny (Mme), professeur de pianos, Seine-
St-Germain, 79.

Didot 1851. page 169

↑ Exemple d'entrée
typique

Texte extrait et entités
nommées associées →



Reco. d'entités nommées (*Named Entity Recognition*)

Identifier et classer des expressions d'un ou plusieurs mots d'un texte en catégories prédéfinies : personne, lieu, organisation, date, prix...

Entités spécifiques dans le cas des annuaires.

Exemple:

PERSON

ACTIVITY

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

LOCATION



CARDINAL

FEATURE TYPE

LOCATION

CARDINAL

Table 1: Entities to recognise in the dataset.

Entity	Count	Description
PER	8788	a person name or a business name, usually referred to as several person names. First names, initials, or civility mentions are included. E.g.: <i>Alibert (Prosper)</i> , <i>Allamand frères et Hersent</i> , <i>Heurtemotte (Vve)</i> .
TITLE	483	an honorary title, either text, glyphs or a combination of the two. E.g: O.  for <i>Officier de la Légion d'Honneur</i> or  for the Great Medal at the London exhibition.
ACT	6472	the profession or social status of a person summarised in the single concept of activity. E.g.: <i>horlogerie</i> , <i>export en tous genres</i> , <i>Conseiller d'État</i> , <i>propriétaire</i> .
LOC	9709	mostly street names (<i>r. de la Jussienne</i>), but may also be neighbourhoods (<i>Marais du Temple</i>) or any indirect spatial references (<i>au Palais Royal</i>).
CARDINAL	8747	a street number, as part of an address (16, 5 bis), or a range of numbers (e.g. 23-25 or 5 à 9).
FT	43	a geographic entity type, used to give more details on a location, e.g. <i>boutique</i> , <i>atelier</i> , <i>fab.</i> or <i>dépôt</i> .

Extraction des données des annuaires

Difficultés : Imperfection des images



Édition & Archivage



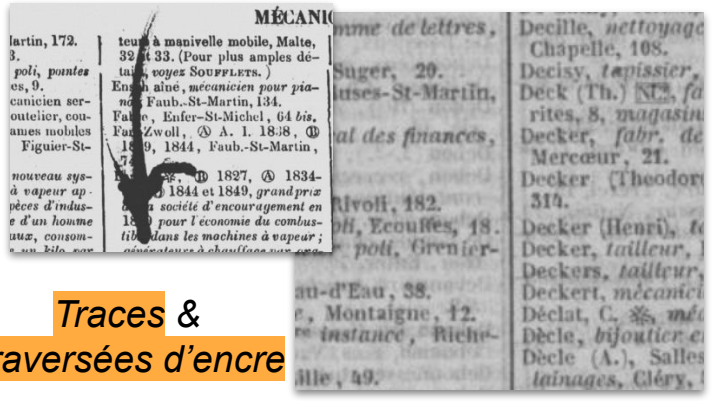
Traces & Traversées d'encre



Numérisation



Compression



Perte des niveaux de gris & Artefacts de compression

Inclinaison & Rotation & Courbure



Difficultés : Résultats OCR bruités

Gavarret ✱, prof. de physique à la faculté de
médecine, Grenelle-St-Germain, 49.


ravarret #, prof. de physique à la faculté de
médecine, Grenelle-St-Germain, 49.

Tesseract v4 output

Daffaut, chaudronnier, r. de la Sourdière,
314

Daflant, c'audronnier, v, de la RARES
Ge OO a x

Tesseract v4 output

 = erreurs

Besoin de **robustesse au bruit OCR**, voire de *post-correction*.

Difficultés : Variation de la structure des entrées

Raison soc.

Activité

Adresse

Durand jeune; pour bas, Charenton, 12 ancien. 18. *

No rue

Prévost-Guillaume, f. ta., r. N.-St.-Mart., 28.

Lefranc Méquignon et co., satins turcs, prunelle, satins et draps de soie, gros de Naples, toiles, coutils, galons, rubans, coulisses et lacets pour chaussures de dames, sommières, flanelles et molletons de soie pour fourrures, r. des Prouvaires, 32.

Planche , R. de Poitou, 9.-H. Armé.

Baronnat frères, soies teintes et écrues, fil-Denis, 257, passage du Renard ; maison à Lyon, r. Cen-

Jamain, orangiste, ⒶS. H.1831, Fossés-St-Marcet, 12.

Appert fils, verres et cristaux, 21-23 Jour.

Mabire, Lourcine, 124.

Besoin d'un système robuste aux **variations** et **bruit syntaxique**, via un **apprentissage** à base d'**exemples**.

Chaîne de traitement implémentée (version fin 2022)

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

Segmentation du canvas
Classification des blocs

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

Transcription

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

Extraction des
informations clés

PERSONNE	ACTIVITÉ	RUE
Lemonnyer,	plomberie,	r. de Bondy,
NUMÉRO	RUE	NUMÉRO
86, et	r. Bouchardon,	1.



Chaîne de traitement implémentée (version fin 2022)

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

Segmentation du canvas
Classification des blocs

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

Transcription

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

Extraction des
informations clés

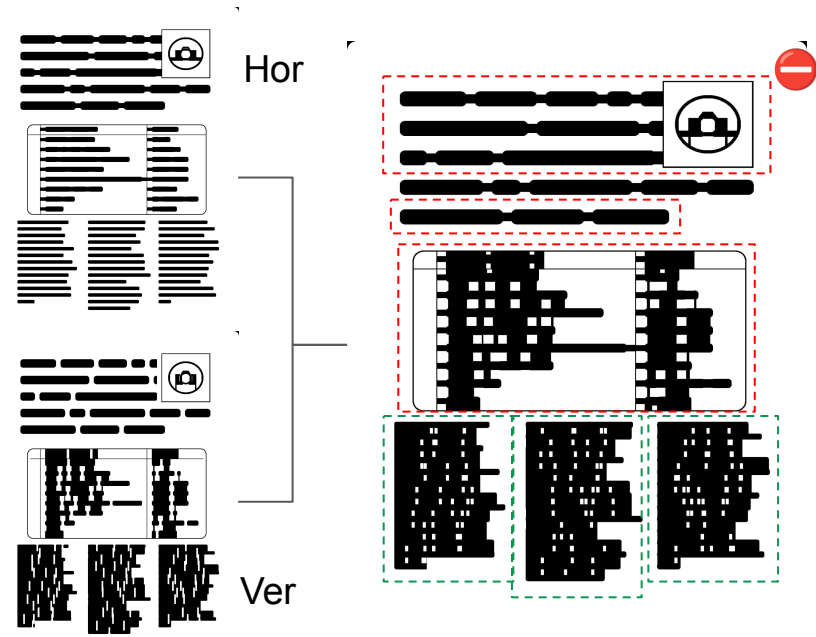
PERSONNE ACTIVITÉ RUE
Lemonnyer, plomberie, r. de Bondy,
NUMÉRO RUE NUMÉRO
86, et r. Bouchardon, 1.



Segmentation du canvas (pour les blocs)

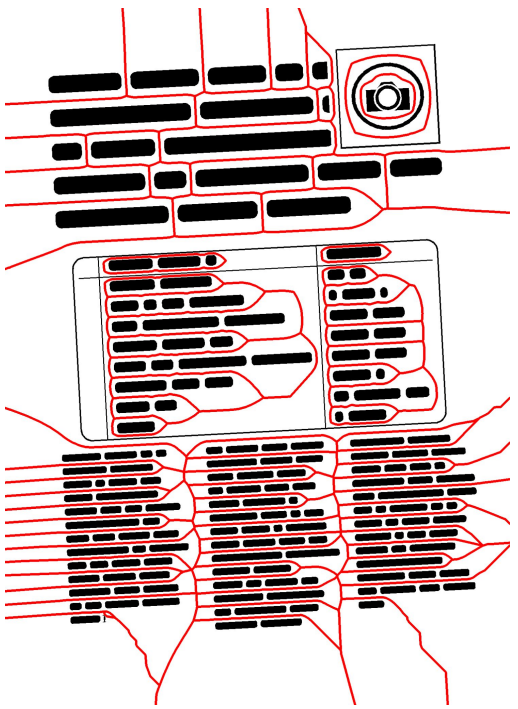
XY-Cut (Projection d'histogramme récursif)

Smearing (ouverture/fermeture morpho)



Segmentation du canvas (pour les lignes)

Ligne de partage des eaux



<p>optique, fait spécialement le compte-fil, microscope à grain, la loupe à réchaud et à botanique, rue des Gravilliers, 3. *</p> <p>Augée, imp. St-Sébastien, 2.</p> <p>Bac (G.), Corderie-du-Temple, 13.</p> <p>Bara, Chaillot, 21.</p> <p>Barat (Vve), cuivre, St-Jean-de-Beauvais, 15.</p> <p>Bardou, Descartes, 38, au fond de la cour, près l'École polytechnique.</p> <p>Batard, successeur de M. Chevalier, tourneur sur métaux pour la petite partie, place du Vieux-Marché-St-Martin, 5 et 7. *</p> <p>Baudry, Charonne, 24.</p> <p>Beaurain, Neuve-Ménilmontant, 8.</p> <p>Beck (L.), tourneur-repousseur, rue de Valenciennes, 10.</p>	<p>rouennais les vieux échange</p> <p>Coursier</p> <p>Eg-St-M</p> <p>Cuvillier</p> <p>David, to de chand et façon glaise de gnoirs, chenets, modèles à ressort entière, 1879, M. Daviron, Eg-St-M</p> <p>Debray, q</p>	<p>TOURNEURS EN MÉTAUX.</p> <p>Adrien, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>André, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>Auguste, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>Benoît, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>Charles, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>Émile, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>François, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>Georges, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>Hippolyte, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>Jacques, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>Jean, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>Jules, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>Louis, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>Martin, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>Paul, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>René, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>Théodore, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>Victor, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>Xavier, tourneur et fabricant de machines, rue de Valenciennes, 10.</p> <p>Zola, tourneur et fabricant de machines, rue de Valenciennes, 10.</p>
--	--	---

Chaîne de traitement implémentée (version fin 2022)

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

Segmentation du canvas
Classification des blocs

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

Transcription

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

Extraction des
informations clés

PERSONNE ACTIVITÉ RUE
Lemonnyer, plomberie, r. de Bondy,
NUMÉRO RUE NUMÉRO
86, et r. Bouchardon, 1.

Reconnaissance du texte (OCR)

Comparaison de plusieurs systèmes OCR sur notre jeu de données “A Dataset of French Trade Directories from the 19th Century (FTD)”

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

Systèmes testés

- [Tesseract v4](#) (v5 pas testée)
- [PERO OCR](#) (code Github + modèle 2020 auteurs)
- [Kraken OCR](#) (modèle générique EN imprimé)

Pas de ré-entraînement.

Bottin 1820
Dufort, bottier, Palais-R., gal. vitrée, 215.
295

Bottin 1827
Baleste, chef aux domaines, S.-Georges, 17.

Bottin 1837
Cattois, pharmac., Bretagne, 46.

Bottin 1854
Fontaine, draperies, Neuve-des-Petits-Champs, 2.

Cambon Almgene 1841
Aron Javal (L.) art. de Paris, r. des Bour-
donnais, 17.

Deflandre 1828
DEVILLERS, r. Croix-des-Pet.-Champs, 25.
Cordonn.

Deflandre 1829
Huguenin, épici., r. de Valois, 8, Pal.-Royal

Didot 1851
**Viéville, fab. de boutons, Aumaire, 48, et place
St-Nicolas-des-Champs, 2.**

Reconnaissance du texte (OCR)

Résultats

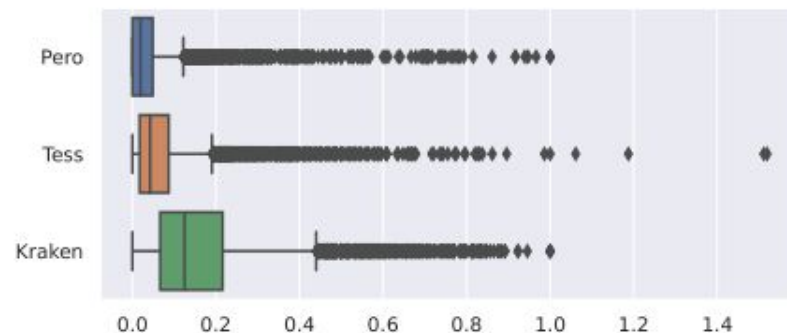
- Très bonne performance de PERO OCR
- Manque de modèles publics pour Kraken

	PERO OCR	Tesseract	Kraken
CER	3.78%	6.56%	15.72%

Autres systèmes libres compétitifs testés depuis

- [Microsoft UniLM/TrOCR](#)
- [Mindee docTR](#)

⇒ OCR plutôt performant, résultats exploitables



Chaîne de traitement implémentée (version fin 2022)

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

Segmentation du canvas
Classification des blocs

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

Transcription

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

Extraction des
informations clés

PERSONNE **ACTIVITÉ** **RUE**
Lemonnyer, plomberie, r. de Bondy,
NUMÉRO **RUE** **NUMÉRO**
86, et r. Bouchardon, 1.

Contribution 1 : *Named Entity Recognition sur texte bruité*

*N. Abadie, E. Carlinet, Joseph Chazalon, B. Duménieu. A Benchmark of Named Entity Recognition Approaches in Historical Documents: Application to 19th Century French Directories. Document Analysis Systems. DAS 2022., May 2022, La Rochelle, France.
<https://github.com/soduco/paper-ner-bench-das22>*

Reconnaissance d'entités nommées à base de réseaux de neurones profonds

1. Quels sont les modèles **récents disponibles** pour réaliser cette tâche de NER ?
2. Ces modèles de NER ont-ils **besoin de beaucoup de données d'entraînement** pour bien fonctionner sur les annuaires ?
3. Ces modèles de NER peuvent-ils produire de **bons résultats malgré le bruit OCR** ?
4. Peut-on **améliorer la résistance de ces modèles de NER au bruit OCR** ?

Modèles NER évalués

Nécessaire pour traiter des entités non standards

Entraînement original

modèle sur étagère

Adaptation de domaine

entraînement spécifique au corpus

SpaCy CNN

Pré-entraînement non supervisé: deep-sequoia
Entraînement supervisé pour le NER: wikiner-fr

~~Pré-entraînement~~
Entraînement supervisé pour le NER: FTD labelled

CamemBERT

Pré-entraînement non supervisé: OSCAR
Entraînement supervisé pour le NER: wikiner-fr

~~Pré-entraînement~~
Entraînement supervisé pour le NER: FTD labelled

CamemBERT *pré-entraîné*

Pré-entraînement non supervisé: OSCAR
Entraînement supervisé pour le NER: wikiner-fr

Pré-entraînement non supervisé: FTD unlabelled
Entraînement supervisé pour le NER: FTD labelled

Les modèles testés sont disponibles sur nos dépôts [HuggingFace](https://huggingface.com) et Zenodo ([10.5281/zenodo.6576008](https://zenodo.org/record/6576008)).

Transcriptions Pero OCR non corrigées pour 845,000 entrées brutes (≈7000 p.)

Transcriptions et annotations manuelles pour 8765 entrées (78 p.)

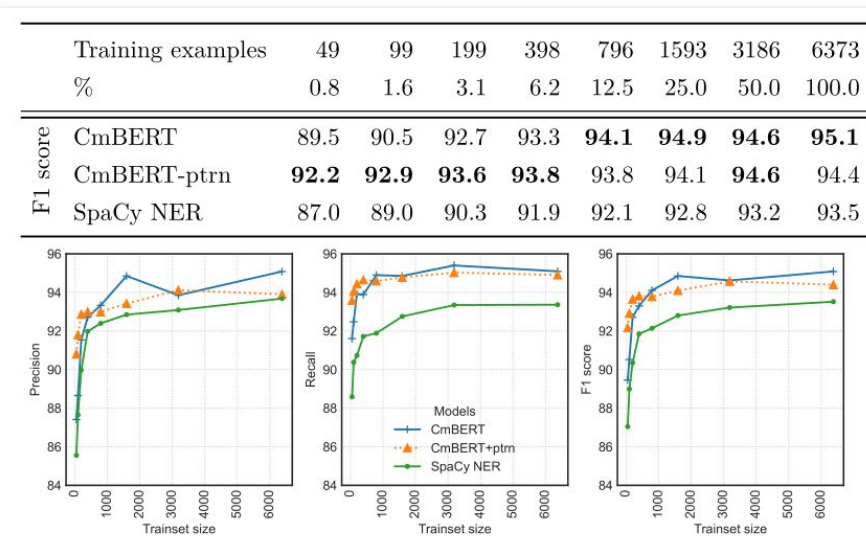
Quantité d'annotations et qualité des résultats

2. A-t-on besoin de beaucoup de données d'entraînement ?

⇒ Non ! En spécialisant des modèles publics, on obtient de **bon résultats avec peu de données d'entraînement**, surtout si on réalise aussi un **pré-entraînement non supervisé**.

3. Les modèles de NER à base de réseaux de neurones profonds sont-ils adaptés pour traiter des annuaires hétérogènes ?

⇒ Oui ! Et les **modèles de type Transformer** obtiennent de très bon résultats.



F1-scores moyens obtenus pour 5 entraînements + tests

Stratégie d'entraînement pour réduire l'influence du bruit OCR sur les résultats

Modèles testés :

- CamemBERT
- CamemBERT pré-entraîné

Entrées :

- Extraits d'annuaires corrigés et annotés manuellement
- Extraits d'annuaires directement produits par OCR (Tesseract V4 et Pero)

Sorties — Prédictions :

- prédictions NER de camemBERT
Avec et sans pré-entraînement

Sorties — Référence :

- Extraits d'annuaires corrigés et annotés manuellement
- Ou annotations manuelles projetées sur les extraits d'annuaires directement produits par OCR (bruités).

Mesures : score F1

Jeux de données utilisés pour entraîner et tester CamemBERT (× 12

variantes)		
Jeux de pré-entraînement	Jeux d'entraînement	Jeux de test
<ul style="list-style-type: none"> • <i>Aucun</i> • PERO (brut) 	<ul style="list-style-type: none"> • Référence • PERO* (* annotations projetées) 	<ul style="list-style-type: none"> • Référence • PERO* • Tesseract* (*annotations projetées)

Extrait du jeu de test de référence:

Annotations manuelles :

Dulay **PER**, chaudronnier **ACT**, r. du Pont- aux Choux **LOC**, 15 **CARDINAL**, 314.

Extrait du jeu de test avec annotations projetées sur du texte bruité :

Dulay **PER**, chaudronnier **ACT**, +. du Pont-anx-Cars **LOC** Ge 7 **CARDINAL** Fe ÊR one

Stratégie d'entraînement pour réduire l'influence du bruit OCR sur les résultats

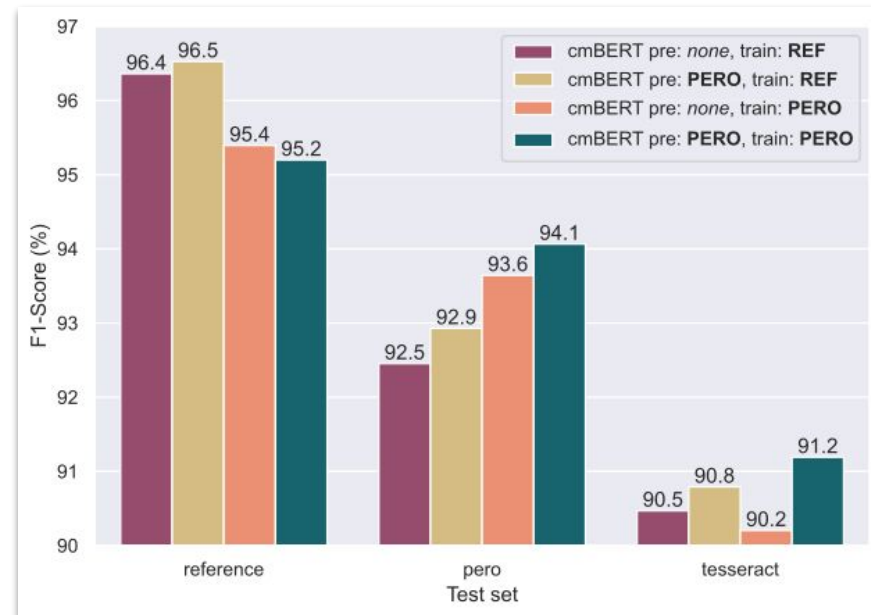
3 bis. Les modèles de NER à base de réseaux de neurones profonds sont-ils adaptés pour traiter des **transcriptions OCR bruitées** ?

⇒ Oui ! Mais les résultats sont moins bons que sur des textes propres.

4. Peut-on rendre les modèles de NER à base de réseaux de neurones profonds **plus robustes au bruit OCR** ?

⇒ Oui ! Il faut les pré-entraîner et les entraîner sur des textes bruités.

⇒ N'entraînez pas votre modèle sur du texte propre si vous voulez traiter des textes OCRisés !



F1-scores moyens obtenus pour 5 entraînements + tests

En pratique...

Pour l'ensemble du corpus :

- 9 821 898 entrées extraites
- 7 260 104 (74%) entrées exploitables, de la forme :

PERSON [**ACT**] **LOC** [**CARDINAL**]

Quelques résultats qualitatifs

1. Cas positifs (ancienne interface)

View 700 Didot 1856a Local Server Handles size Force compute

Boulon, Charenton, 81.

Bouquet, M. H. Exposition universelle de 1855, breveté s. g. d. g., spécialité de tables de salles à manger, nouveau système mécanique donnant la facilité à une seule personne de pouvoir l'ouvrir, n'importe sa grandeur, Faub.-St-Antoine, 99, passage du Bras-d'Or. *

Bourdier, fab. de divans en tous genres, fait tout ce qui concerne sa partie, Charenton, 16.

Bourdier aîné, Nve-de-Lappe, 16.

Bourgade. Rumford. 14.

(0/159 entries checked)

ENTRY

OCR NER Form JSON

PER ACT LOC CARDINAL FT TITRE

↑ CLEAR

Boulon, Charenton, 81.

Checked

Comment

• EAUX AUROPHILE ET ARGENTO-PHILE Berger, eaux spéciales pour nettoyages, déposées conformément à la loi. L'Aurophile, liquide pour nettoyer avec la plus grande facilité, propreté et sans altération, toutes pièces en bronze dorés, mates ou polies, montées ou non, avec mar-bre, etc., comme pendules, lustres, candélabres, garnitures de meubles, les statuettes en biscuit, marbre, etc., les fleurs ou autres sujets très-compliqués en porcelaine, les camées, l'ivoire, la nacre, les cristaux polis ou dépolis, et avec beaucoup de précaution, les dorures sur bois. L'Argentophile, liquide pour nettoyer avec la plus grande facilité et sans altération toutes pièces d'orfèvrerie mates ou brunies, soit argent, argentées ou plaquées : la joaillerie, la bijouterie, ainsi que

View 700 Didot 1856a Local Server Handles size Force compute

Bouquet, M. H. Exposition universelle de 1855, breveté s. g. d. g., spécialité de tables de salles à mangèr, nouveau système mécanique donnant la facilité à une seule personne de pouvoir l'ouvrir, n'importe sa grandeur, Faub.-St-Antoine, 99, passage du Bras-d'Or. *

Bourdier, fab. de divans en tous genres, fait tout ce qui concerne sa partie, Charenton, 16.

Bourdier aîné, Nve-de-Lappe, 16.

Bourgade, Rumford, 14.

Bourgoin. St-Antoine. 176.

(0/159 entries checked) ENTRY

OCR NER Form JSON

PER ACT LOC CARDINAL FT TITRE

↑ CLEAR

Bouquet, M. H. Exposition universelle de 1855, breveté s. g. d. g., spécialité de tables de salles à mangèr, nouveau système mécanique donnant la facilité à une seule personne de pouvoir l'ouvrir, n'importe sa grandeur, Faub.-St-Antoine, 99, passage

Checked

Comment

• EAUX AUROPHILE ET ARGENTO-PHILE Berger, eaux spéciales pour nettoyages, déposées conformément à la loi. L'Aurophile, liquide pour nettoyer avec la plus grande facilité, propreté et sans altération, toutes pièces en bronze dorés, mates ou polies, montées ou non, avec marbre, etc., comme pendules, lustres, candélabres, garnitures de meubles, les statuettes en biscuit, marbre, etc., les fleurs ou autres sujets très-compliqués en porcelaine, les camées, l'ivoire, la nacre, les cristaux polis ou dépolis, et avec beaucoup de précaution, les dures sur bois. L'Argentophile, liquide pour nettoyer avec la plus grande facilité et sans altération toutes pièces d'orfèvrerie mates ou bruniées, soit argent, argentées ou plaquées : la joaillerie, la bijouterie, ainsi que

View 700 Didot 1856a Local Server Handles size Force compute

Bourdier, fab. de divans en tous genres, fait tout ce qui concerne sa partie, Charenton, 16.

Bourdier aîné, Nve-de-Lappe, 16.

Bourgade, Rumford, 14.

Bourgoin, St-Antoine, 176.

Bourlier, Charonne, 99.

Boutard, Traversière, 76.

*Boutung (seule maison connue depuis trente ans), fab. et magasin de meubles en tous genres, modernes, antiques et fantaisie; tient aussi le siège; commission et exportation; Faub.-St-Antoine, 97: ci-devant même rue. 23. **

(0/159 entries checked)

ENTRY

OCR NER Form JSON

PER ACT LOC CARDINAL FT TITRE

CLEAR

Bourdier, fab. de divans en tous genres, fait tout ce qui concerne sa partie, Charenton, 16.

Checked

Comment

- EAUX AUROPHILE ET ARGENTO-PHILE Berger, eaux spéciales pour nettoyages, déposées conformément à la loi. L'Aurophile, liquide pour nettoyer avec la plus grande facilité, propreté et sans altération, toutes pièces en bronze dorés, mates ou polies, montées ou non, avec marbre, etc., comme pendules, lustres, candélabres, garnitures de meubles, les statuettes en biscuit, marbre, etc., les fleurs ou autres sujets très-compliqués en porcelaine, les camées, l'ivoire, la nacre, les cristaux polis ou dépolis, et avec beaucoup de précaution, les dures sur bois. L'Argentophile, liquide pour nettoyer avec la plus grande facilité et sans altération toutes pièces d'orfèvrerie mates ou brunies, soit argent, argentées ou plaquées: la joaillerie, la bijouterie, ainsi que

View 700 Didot 1856a

Local

Server

Handles size

Force compute

Bourdier aîné, Nve-de-Lappe, 16.
 Bourgade, Rumford, 14.
 Bourgoïn, St-Antoine, 176.
 Bourlier, Charonne, 99.
 Boutard, Traversière, 76.
 Boutang (seule maison connue de-
 puis trente ans), *fab. et magasin*
de meubles en tous genres, mo-
dernes, antiques et fantaisie;
tient aussi le siège; commission et
exportation; Faub.-St-Antoine,
*97; ci-devant même rue, 23. **

(0/159 entries checked)

ENTRY

OCR NER Form JSON

PER

ACT

LOC

CARDINAL

FT

TITRE

CLEAR

Bourdier aîné, Nve-de-Lappe, 16.

 Checked

Comment

• **EAUX AUROPHILE ET ARGENTO-PHILE** Berger, eaux spéciales pour nettoyages, déposées con- formément à la loi. L'Aurophile, liquide pour nettoyer avec la plus grande facilité, propreté et sans altération, toutes pièces en bronze dorés, mates ou polies, montées ou non, avec mar- bre, etc., comme pendules, lus- tres, candélabres, garnitures de meubles, les statuettes en bis- cuit, marbre, etc., les fleurs ou autres sujets très-compli- qués en porcelaine, les camées, l'ivoire, la nacre, les cristaux po- lis ou dépolis, et avec beaucoup de précaution, les dorures sur bois. L'Argentophile, liquide pour nettoyer avec la plus grande fa- cilité et sans altération toutes pièces d'orfèvrerie mates ou brunies, soit argent, argentées ou plaquées : la joaillerie, la bi- jouterie, ainsi que

Quelques résultats qualitatifs
2. Cas limites (nouvelle interface)

Didot 1851a, vue 800

Figueran, Enghien, 7.

Figueran PER , Enghien LOC , 7 CARDINAL .

Cas typique : entrée courte avec OCR et NER fonctionnels.

Didot 1851a, vue 700

Berendorff (J.), M. H. 1844, Mouf-
fetard, 294, *atelier de construction
de machines à vapeur et autres*,
Ⓐ S. E., Ⓐ 1844, Ⓞ 1849.

Berendorff(J.), PER M. H. TITRE 18 44 TITRE , Mouf-
fetard LOC , 294 CARDINAL , atelier de construction de
machines à vapeur et autres ACT , S. TITRE E., O 1844, O
1849.

Cas typique : entrée longue avec entités principales (PER, LOC, CARD.) bien reconnues malgré l'inclusion de (“,”), et des erreurs NER sur les autres entités.

Duverneuil et La Tyna, 1806

Benaed, R. des Bons-Enfans, 19.

Benaed PER , R. des Bons-Enfans LOC , 19 CARDINAL .

Erreur OCR sur le nom ⇒ cible pour la correction par redondance

**Chaussard, R. de Grenelle, Halle-aux
Blés, 33.**

Chaussard PER , R. de Grenelle, Halle-aux Blés LOC , 33 CARDINAL .

Exemple positif, mais fusion de plusieurs niveaux de localisation

Deflandre 1829, vue 500

NOEL aîné et fils. Maison de commerce pour la lithographie; r. N.-des-Petits-Champs , galerie Colbert, esc. A.

NOEL aîné et fils PER . Maison de commerce pour la lithographie ACT ; r. N.-des-Petits-Champs , galerie Colbert LOC . esc. A

Exemple positif, mais fusion de plusieurs niveaux de localisation

Contribution 2 : NER hiérarchique et modèles d'annotation

Stage de M2 de **Solenn Tual** (ENSG / IGN) encadrée par N. Abadie, B. Duménieu et J. Chazalon.
Preprint article : <https://hal.science/hal-03994759>

Motivation



Problèmes attaqués :

1. Comment donner **plus de structure** aux données extraites (ex : adresse = rue + no) ?
2. Peut-on tirer profit d'une **connaissance hiérarchique** pour améliorer les prédictions "à plat" ?

Approche :

- **Apprendre** avec des données hiérarchique
- **Produire** une représentation hiérarchique

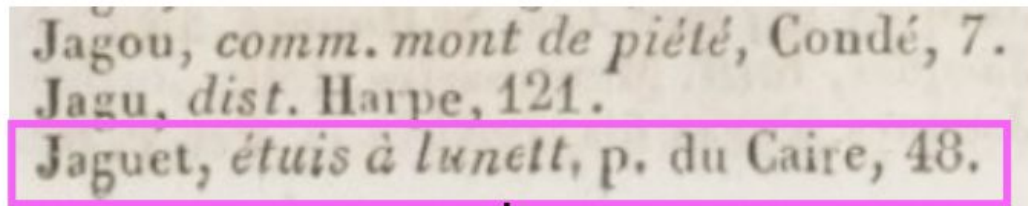
3. Peut-on améliorer la séparation entre les **entités voisines** ?

Approche :

- Tester d'autres **modèles d'annotation** du texte (*des tokens*)

Représentation hiérarchique

Historical source



OCR output

Jaguet, étuis à lunett, p. du Caire, 48.

Nested NER output



Modèles d'annotation : IO vs IOB2

Token	IO	IOB2
Aubery	I-PER	B-PER
je	I-PER	I-PER
.	I-PER	I-PER
r	I-LOC	B-LOC
.	I-LOC	I-LOC
Quincamp	I-LOC	I-LOC
.	I-LOC	I-LOC
pass	I-LOC	B-LOC
.	I-LOC	I-LOC
Beaufort	I-LOC	I-LOC
.	O	O



Différence :
Annotation de la frontière ou non.



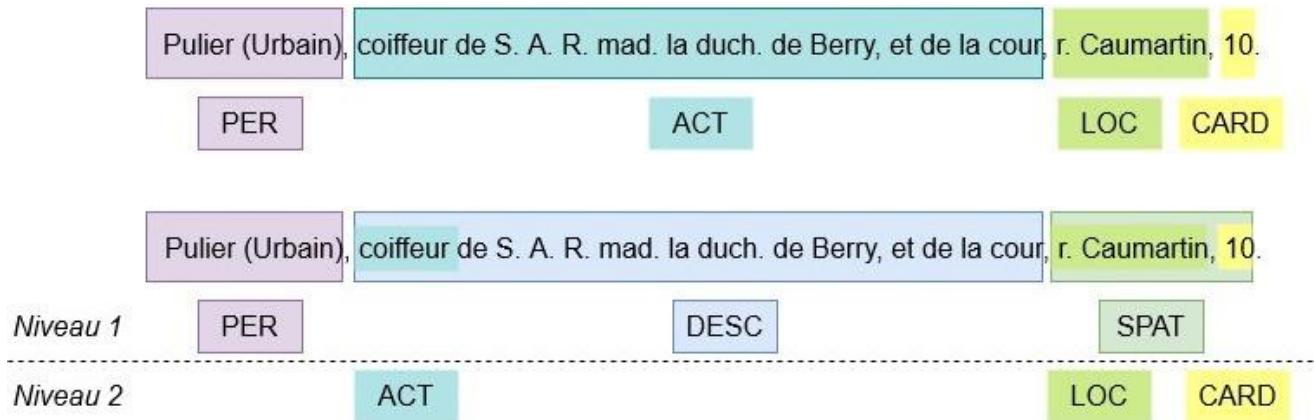
Impact important :
Augmentation du nombre de classes (×2) donc diminution du nombre d'exemples par cas !



Méthodes testées

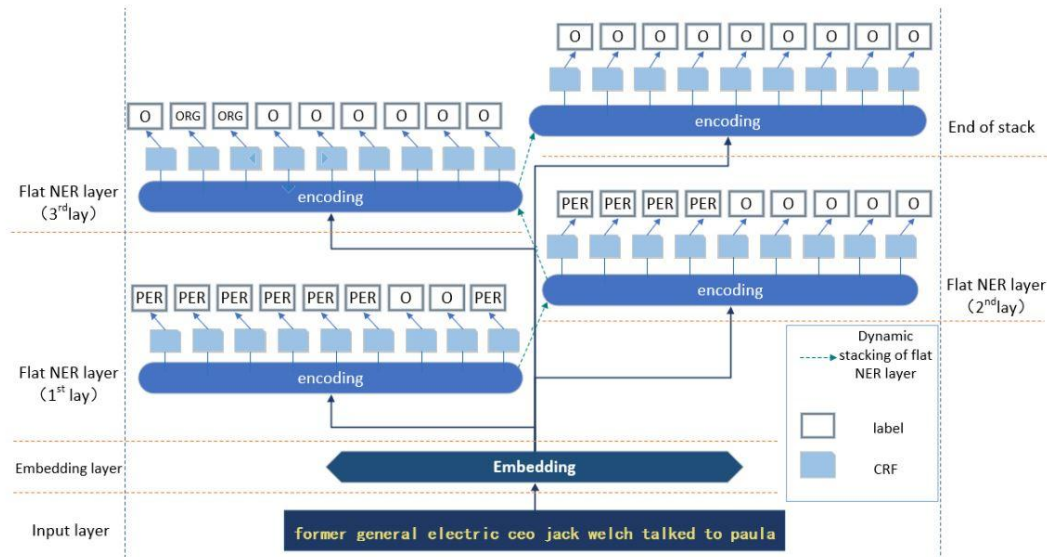
[M1] Prédiction de chaque niveau indépendamment

- (+) Entraînement simplifié (réseaux simples, données simples)
- (-) 2 réseaux (ou 2 têtes a minima) donc plus de paramètres et de calculs
- (-) Pas de garantie de cohérence entre niveaux



[M1] adaptée de : Jia, L., Liu, S., Wei, F., Kong, B., Wang, G.: *Nested Named Entity Recognition via an Independent-Layered Pretrained Model*. IEEE Access 9, 109693–109703 (2021).

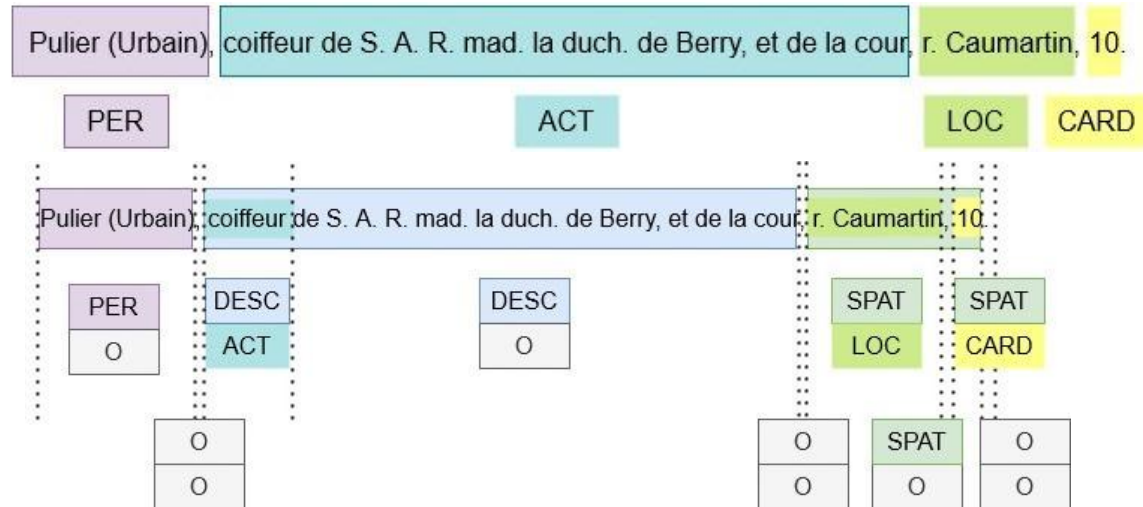
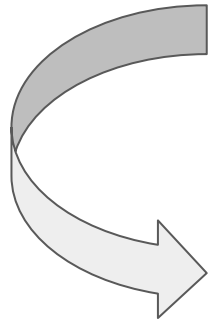
<https://doi.org/10.1109/ACCESS.2021.3102685>



Méthodes testées

[M2] Prédiction conjointe

- (+) Garantie de cohérence : seul les combinaisons légales peuvent être prédites
- (+) 1 seul réseau à entraîner
- (-) Annotation plus complexe avec une augmentation du nombre de labels

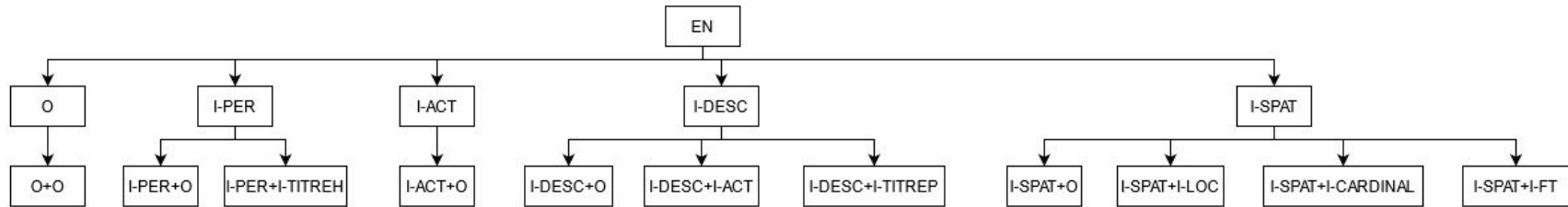


[M2] adaptée de : Agrawal, A., Tripathi, S., Vardhan, M., Sihag, V., Choudhary, G., Drag- oni, N.: *BERT-Based Transfer-Learning Approach for Nested Named-Entity Recognition Using Joint Labeling*. Applied Sciences 12(3), 976 (2022). <https://doi.org/10.3390/app12030976>

Méthodes testées

[M3] Prédiction conjointe avec fonction de coût hiérarchique

Mêmes propriétés que [M2], mais avec une pénalité plus faible pour les erreurs entre des labels proches.



[M3] adaptée de : Bertinetto, L., Mueller, R., Samangoei, S., Lord, N.A.: *Making Better Mistakes: Leveraging Class Hierarchies with Deep Networks*. CoRR abs/1912.09393 (2019).

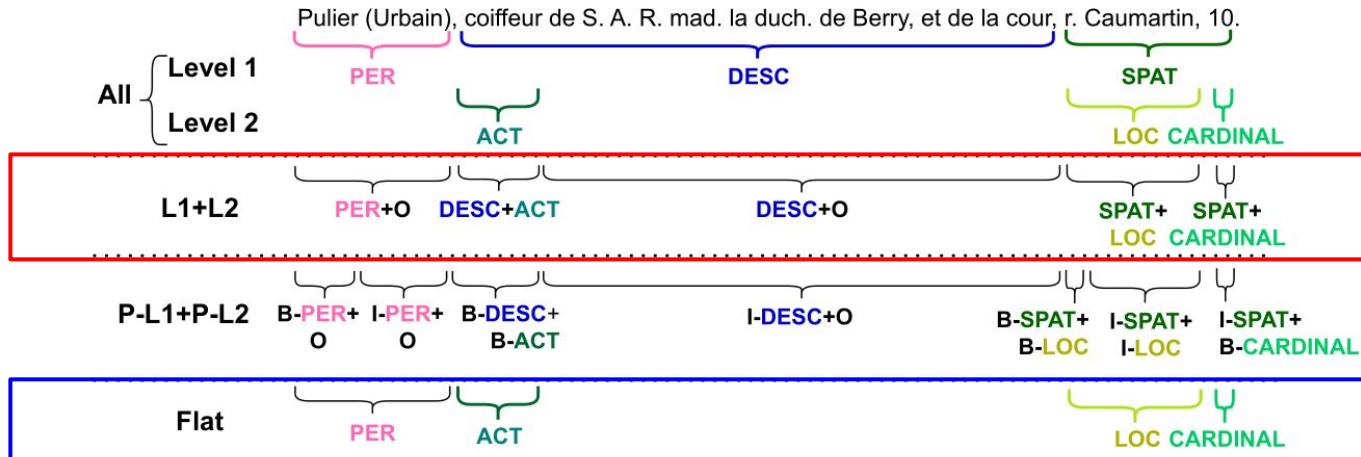
<https://doi.org/10.48550/arXiv.1912.09393>

Protocole expérimental

Mêmes données d'annuaires, avec des **annotations enrichies**.

Évaluation stricte des combinaisons (exactement les mêmes labels, exactement aux mêmes positions), au niveau entité.

Métriques :



Peut-on extraire des données structurées hiérarchiques ?

Les annotations hiérarchiques permettent-elles de meilleurs résultats en "flat" ?

Résultats sur données bruitées réelles (moyenne 5 runs)

IO > IOB2
en général

Model and tags	All	L1+L2	L1	L2	P-L1+P-L2	Flat	
M1	CmBERT IO	93.8	93.4	93.1	94.6	93.4	94.2
	CmBERT IOB2	93.5	92.9	93.1	94.0	93.1	92.7
	CmBERT+ptrn IO	94.3	93.8	94.1	94.5	93.8	94.4
	CmBERT+ptrn IOB2	94.1	93.5	93.7	94.5	93.7	94.5
M2	CmBERT IO	93.8	94.1	93.3	94.4	94.1	94.5
	CmBERT IOB2	93.8	94.2	93.2	94.5	94.3	94.7
	CmBERT+ptrn IO	93.9	94.1	93.4	94.4	94.1	94.6
	CmBERT+ptrn IOB2	93.7	94.1	93.1	94.5	94.2	94.8
M3	CmBERT IO	94.1	94.4	93.5	94.8	94.4	94.8
	CmBERT IOB2	93.5	93.9	92.9	94.3	94.0	94.6
	CmBERT+ptrn IO	94.1	94.4	93.6	94.8	94.4	94.9
	CmBERT+ptrn IOB2	93.7	94.0	93.1	94.4	94.2	94.8

Peut-on extraire des
données structurées
hiérarchiques ? → **Oui**

Les annotations hiérarchiques
permettent-elles de meilleurs
résultats en "flat" ? → ...

Résultats du modèle “flat” : idem modèle structuré

Dataset Test		F1-Score (in %)
OCR	CmBERT	94.7
	CmBERT+ptrn	94.9

Les annotations hiérarchiques permettent-elles de meilleurs résultats en “flat” ?
Pas dans ce cas.

Méthode et modèle d'annotation à privilégier

[M3] Prédiction **conjointe** avec fonction de **coût hiérarchique** :

- relativement simple à mettre en œuvre
- n'introduit qu'un surcoût faible, si le nombre de classes n'est pas trop important
- mais pas de gain par rapport au modèle de base pour des entités "à plat"

Annotations IO

- évite de doubler le nombre de classes, dans la mesure où les frontières entre entités ne semblent pas le problème essentiel, et dans un contexte avec peu de données pour spécialiser l'approche.

Contribution 3 :
Modèle de langage mixte visuel/textuel
pour la séparation d'entrées

Travaux de N. Abadie, B. Duménieu, E. Carlinet et J. Chazalon.

Preprint article : <http://arxiv.org/abs/2302.08948>

Motivation et approche

Problèmes attaqués :

1. Comment tirer profit à la fois de l'**information visuelle et textuelle** pour mieux **séparer les entrées** ?
2. Comment s'affranchir des **sauts de colonne** et de **pages** ?
3. Est-il possible de **séparer les entrées** et faire le **NER simultanément** ?

Opportunité : documents majoritairement textuels, ordre de lecture simple

Motivation et approche

Problèmes attaqués :

1. Comment tirer profit à la fois de l'**information visuelle et textuelle** pour mieux **séparer les entrées** ?
2. Comment s'affranchir des **sauts de colonne** et de **pages** ?
3. Est-il possible de **séparer les entrées** et faire le **NER simultanément** ?

Opportunité : documents majoritairement textuels, ordre de lecture simple

Solution : un modèle de langage mixte **visuel** et **textuel**

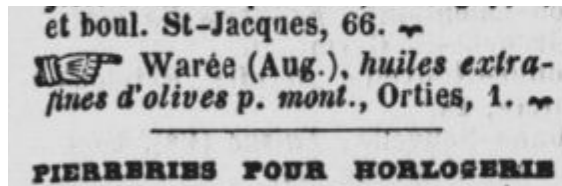


Image originale

```
<page_break><small_lhspace> et boul. St-Jacques, 66.<glyph_bat>  
<large_rhspace> <line_break> <small_lhspace> Warée (Aug.), <it>huiles  
extra- <small_rhspace> <line_break> <small_lhspace> fines d'olives p.  
mont.</it>, Orties, 1. <glyph_bat> <small_rhspace> <line_break>  
<large_lhspace> <hrule> <large_rhspace> <line_break> <small_lhspace>  
<b>PIERRERIES POUR HORLOGERIE </b> <small_rhspace>
```

Nouvel input pour le système "NER+sép. entrées"

Tout projeter dans le référentiel textuel

Possible grâce à :

- la flexibilité des outils existants (tokens personnalisés)
- l'information de position des blocs et lignes détectés

Permet d'utiliser le même réseau que le NER, éventuellement simultanément !

- Entrée : nouveau flux augmenté
- Sortie : Il suffit de prédire des points de coupure, en plus des marqueurs de classe EN.

Nouveaux tokens visuels proposés, injectés dans le flux de tokens textuels :

- Tokens structurels : sauts de ligne, colonne, page, espaces horiz. et vert.
- Tokens graphiques : glyphes hors charset OCR, lignes, images...
- Tokens de présentation : gras, italique, taille relative texte...

Expériences

1. Séparation d'entrée seulement

1.1 → 1.6 Texte seul, puis ajout `<breaks>`, puis ajout `<spaces>` (variantes)

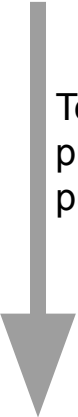
1.7 Tokens visuels seulement (texte remplacé par `<texte_masqué>`)

Résultats (triés par F-score de séparation d'entrées)

1. Comment tirer profit à la fois de l'information visuelle et textuelle pour mieux séparer les entrées ?

⇒ Injecter de nouveaux tokens, et profiter d'une amélioration continue !

Text,
puis breaks,
puis espaces



Text + Visuel (best)

Texte seul

Visuel seul

# experiment	Experimental setup					EBEGIN & EEND (%)			Overall NER F-score
	text	ner on text content	visual breaks	Left spaces	right spaces	Precision	Recall	F-score	
1.1	yes	no	no	no	no	97.3	95.5	96.4	-
1.2	yes	no	yes	no	no	99.5	98.0	98.7	-
1.3	yes	no	yes	abs.	no	99.7	98.3	99.0	-
1.4	yes	no	yes	rel.	no	99.7	98.2	98.9	-
1.5	yes	no	yes	abs.	abs.	99.6	98.5	99.0	-
1.6	yes	no	yes	rel.	abs.	99.4	98.6	99.0	-
1.7	no	no	yes	rel.	abs.	98.4	97.9	98.2	-

Expériences

1. Séparation d'entrée seulement

1.1 → 1.6 Texte seul, puis ajout <breaks>, puis ajout <spaces> (variantes)

1.7 Tokens visuels seulement (texte remplacé par <texte_masqué>)

2. Séparation d'entrée et NER simultané

2.1 Labels “flat” et utilisation des tokens d'espaces pour marquer les séparations

2.2 Labels joints (E_BEGIN+PER, E_BEGIN+ACT...)

Résultats (triés par F-score de séparation d'entrées)

2. Est-il possible de séparer les entrées et faire la REN *simultanément* ?

⇒ **Oui ! Et avec assez de données** d'entraînement (ou pas trop de classes), on **améliore les performances** de la séparation d'entrées.

	# experiment	Experimental setup					EBEGIN & EEND (%)			Overall NER F-score
		text	ner on text content	visual breaks	Left spaces	right spaces	Precision	Recall	F-score	
Joint labels →	2.2	yes	yes	yes	rel.	abs.	94.7	56.3	70.6	93.0
Texte seul →	1.1	yes	no	no	no	no	97.3	95.5	96.4	-
Visuel seul →	1.7	no	no	yes	rel.	abs.	98.4	97.9	98.2	-
	1.2	yes	no	yes	no	no	99.5	98.0	98.7	-
	1.4	yes	no	yes	rel.	no	99.7	98.2	98.9	-
	1.3	yes	no	yes	abs.	no	99.7	98.3	99.0	-
	1.6	yes	no	yes	rel.	abs.	99.4	98.6	99.0	-
Text + Visuel (best) →	1.5	yes	no	yes	abs.	abs.	99.6	98.5	99.0	-
Flat labels →	2.1	yes	yes	yes	rel.	abs.	99.3	99.0	99.2	94.0

Plusieurs bonus

*2. Comment s'affranchir des **sauts de colonne et de pages** ?*

⇒ Traiter le **flux de texte en continu** ! Il n'y a **plus de limite** de colonne ou de page.

*Est-ce plus **coûteux** ?*

⇒ **Non, ni en entraînement ni en inférence** !
Le même modèle réalise 2 tâches avec un surcoût négligeable.

⇒ De plus, les données sont **rapides à annoter** (pour la séparation d'entrées).

*Est-ce **fiable** ?*

⇒ On tire profit du **meilleur de l'étage d'analyse de layout** (détection de macro blocs et de leur ordre) **et de l'analyse de texte** en intégrant **TOUTE l'information** utile.

*Et par rapport à une approche **end-to-end** ?*

⇒ L'entraînement est **très rapide et simple**, et il est possible de venir **corriger les éléments détectés** et **relancer les traitements** en mode **interactif**.

Outils existants et ouverture en cours

Assister le travail humain avec des outils automatiques



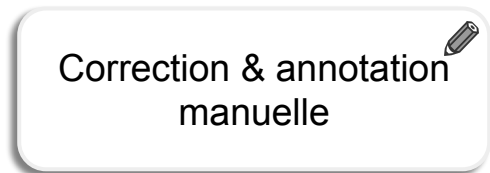
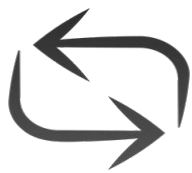
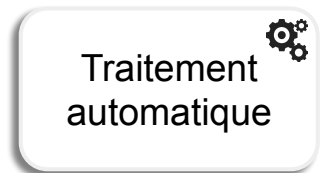
API IIIF
E.g {BNF Gallica



Export structuré
JSON / TEI XML

Interopérabilité

Des micro-services intégrés à l'annotation semi-automatique



D'autres avantages:

- App web (distante ou locale) facile à déployer
- Des services de traitements externalisés (on peut changer et combiner des méthodes)

Cabanis et Cie PER , association pour la construction des machines à éle- ver les eaux, pompes domestiques, moteurs hydrauliques de toute na- ture, machines soufflantes pour les forges, machines à eaux de Sellz, presses à découper, outillage et réparation de machines ACT , Vinai- griers LOC , 32 CARDINAL .

Comment Tags

Checked

SAVE

MÉCANICIENS.

697

lons ronds et demi-ronds et carvés, et de toute sorte de moulure ; spécialité de fabrication de broches en acier et cuivre pour les peignes à tisser draps, voiles et ouvertures, toile métallique pour tout ce qui concerne le tissage, M. H. 1849, Amandiers-Popincourt, 19.

Auger (Vve), mécanicien, fabr. d'importe- pièces pour festons de garnitures de robes et de mantelets, importe- pièces pour cartonnage, semelles de souliers, etc., tient un assortiment de maillets, billets et plombs, pass. de la Trinité, 77.

Avoyne-Bainée, serrurier-mécan., fab. de lits en f. r., 1839-14, 1819, Boulangers-St-Victor, 22.

Bijard, quai Jemmapes, 248.

Baptrosse pere, tout ce qui a rapport à la fabrique d'indennes, Charenton, 58.

Barbot, Popincourt, 58.

Barbès, Trois-Bornes, 21.

Baridou, Descartes, 38.

Bariquand, horloger-mécanicien, entreprend toute espèce de fabrication nouvelle, quelle que soit la précision demandée, et construit toutes sortes de machines et outils pour Messieurs les inventeurs; vis cylindriques et coniques, découpage à façon, fendage de roues; pièces détachées fabriquées par procédés mécaniques pour toute espèce d'industrie, 1849, Ecoiffes, 20. au Mar., près le marché des Blancs-Manteaux. *

Baudat, constructeur de mécaniques à scier le placage. Le bois on

Becker, mécanicien-graveur, breveté d'invention (sans garantie du gouvernement), fabrique spéciale de presses à copier, presses à timbre sec et bumide, presses à cacheter, presses à perçusion pour l'extraction des matières végétales et pour le satinage des papiers, composteurs pour papiers à lettres, pour raisons de commerce, griffes de toutes sortes, et généralement toute la gravure sur métaux, St-Denis, 380, passage Lemoine. *

Beerstecher (A.), Lourcine, 88.

Berendorff (J.), M. H. 1844, Moulletard, 294, atelier de construction de machines à vapeur et autres, 1849, S. E., 1844, 1849.

Bergeron, 1849, Marais-du-Temple, 73.

Bernier aîné, outils établis, presses, 1844, Faubourg-St-Antoine, 89.

Bernier, Menilmontant, 90.

Berthet, Simon-le-Franc, 13.

Berton, Neuve-St-Denis, 12.

Bertrand, mécanicien, breveté (sans garantie du gouvernement), construit toutes espèces de machines et outils, spécialité de machines pour les fabricants de matières premières de chapellerie, machines à vapeur et tours pour les fabricants de chapeaux; tondeuses et souffleuses de poils de lièvres et de lapins, réparation et entretien de machines à vapeur; on trouve toujours chez lui des machines à vapeur toutes prêtes à être mises en place, Vieille-du-Temple, 58, anc. 72

Boslay (Ch.) 1839 1849

province, à des prix très-moderés, St-Jacques, 261. *

Bosche aîné, ingénieur-mécanicien, inventeur de divers perfectionnements au métier Jacquard, S. E. 1850, Amandiers-Popincourt, 22, ci-devant St-Maur, 14. *

Boscher, Bondy, 70.

Bosquillon, 1823-27-34, constructeur de perçages accélérés et de mécaniques Jacquard parisiennes, Paradis-Poissonnière, 20.

Boucher, Saint-Pierre-Popincourt, 18.

Bouchon, dépôt de moulins à bras portatifs, Nve-St-Nicolas, 16.

Boulley (Ene), mécanicien, fab. de machines à couper les peaux de lapins, souffleuses de toutes dimensions marchant à bras et à la vapeur, tours de chapeliers, 1849, Francs-Bourgeois-Marais, 3, ci-devant Beaubourg, 59. *

Bouhon, appareil dit cale à léau, breveté sans garantie du gouvernement, destiné au soulagement des chevaux dans les montées, 1849, Société d'encouragement 1850, pl. Dauphine, 71. *

Bourdier, bandagiste-herniaire, mécanicien-orthopédiste, fournisseur de la Société protestante de prévoyance et de secours mutuels de Paris, fabrique spéciale de toutes espèces de ressorts et de bandages, béquilles, appareils orthopédiques, ceintures hypogastriques, bas lacés en tous genres, suspensoirs, objets d'allaitement, et tous articles en gomme élastique, Cadran 5.

gnier, mécanique nouvelle, presse à copier les lettres, perfectionnée 1849, Saint-Jacques, 169, près le Panthéon. *

Brocard (Léon), machines à papier, St-Gilles, 12.

Burckel, fab. étrilles, fuseaux et broches, Nve-de-Lappe, 3.

Busser, fab. de presses lithographiques, machines à réparer, fait les réparations à des prix modérés, Lamartine, 18-20, ci-devant Coquenard. *

Butt, Buisson-St-Louis, 10.

Cabanis et Cie, association pour la construction des machines à élever les eaux, pompes domestiques, moteurs hydrauliques de toute nature, machines soufflantes pour les forges, machines à eaux de Sellz, presses à découper, outillage et réparation de machines, Vinai-griers, 32.

Chouet, breveté (sans garantie du gouvern.) pour les porte-moules à bougies emboutis d'un seul morceau, et les moules montés sans aucune soudure, et pouvant les monter et les démonter soi-même; fabricant de mandrins, fondeur de moules de chandelles, bougies, cierges, spécialité, 1849, Marché-aux-Vaux, 4.

Calamel, Dupetit-Thouars, 10.

Calard (T.-F.), fab. de feuilles métalliques percées en fer, cuivre et zinc; tôle en râpe pour le nettoyage des blés, cribles à main pour remplacer la peau.—Atelier de construction de machines à percer, à cintrer les fers à froid et à ajuster les ressorts de voitures. Leclerc. 8. Faub.-St-Jac-

Phases d'ouverture des outils

Phase 0 : partage en **interne**, validation de l'utilité dans le cadre spécifique de SoDUCo, publication de prototypes isolés (recherche reproductible)

Phase 1 : tests avec d'**autres projets pilotes**, identification des éléments les plus utiles sans adaptation réelle (cas du projet AGODA au BnF Datalab)

Phase 2 : **adaptation à d'autres projets similaires** et choix d'une solution technique pour une évolution à long terme

Phase 3 : **ouverture complète** avec un soin particulier apporté à l'interopérabilité, la documentation, la maintenance, la sécurité, etc. en actant un positionnement et une démarche centrée sur quelques usages

Perspectives — Phase 2 d'ouverture en cours

Encore une plateforme d'annotation / reconnaissance de documents imprimés ?

⇒ Non ! Vers une **collection** d'outils **ouverts**, **minimaux**, laissant les historiens **constituer leurs corpus** et y “superposer” **leurs données**.

Nos objectifs à venir :

- **Consolider notre interface** de visualisation et d'annotation (page web unique, sans besoins d'hébergement ni d'installation)
- **Déployer des services** pour une utilisation à la demande sans nécessiter une machine puissante (OCR, NER, etc. — un hébergement HumaNum ?)
- **Enrichir notre boîte à outils** (correction de courbure des pages, export TEI, intégration IIIF...)
- **Partager** plus largement ces outils (*open source*, documentation, voire formation)

En résumé

À retenir à propos de SoDUCo

1. Des plans et annuaires à une analyse de l'évolution de Paris au XIXe
2. Sur les contenus textuels, une chaîne d'extraction pragmatique et efficace
3. Recherche ouverte et reproductible (prototypes disponibles)
4. Outils ouverts ou en cours d'ouverture
5. Données ouvertes ou en cours d'ouverture
6. Un [cycles de séminaires](#) pour continuer la réflexion collective

