

Les données géohistoriques produites au cours du programme SoDUCo. Évaluation d'une « source numérique »

Pascal Cristofoli, Julie Gravier, Stéphane Baciocchi, Bertrand Duménieu, Solenn Tual, Nathalie Abadie

4e Séance SoDUCo/BnF - séminaire de clôture de l'ANR SoDUCo, mardi 7 novembre

RES(T)ITUER LES ANNUAIRES COMMERCIAUX ET LES ÉVOLUTIONS DE L'ESPACE PARISIEN DU XIX^E SIÈCLE

Lundi 6 novembre 2023 (9h30-17h30) en salle 70 – Buffet repas au Belvédère (12h15-13h30)

09h30. M. Carlin (BnF DataLab) : *Ouverture de la journée & SoDUCo team, Le programme de recherche SoDUCo. Bilan*

10h00-12h15. **SESSION** « MÉTADONNÉES, CATALOGAGE & DIFFUSION DE DONNÉES » avec J. Schoonman, B. Spaan et M. Claeys Bouuaert (Allmaps), N. Vigneron (Wikimedia France) et A. Joffres (Huma-num)

14h00-14h45. **KEYNOTE** - M. Barthelemy (CEA) et J. Gravier (EHESS) : « Lois d'échelle » et dynamiques des activités urbaines dans une ville en croissance.

15h00-17h30. **SESSION** « EXTRACTION & STRUCTURATION DE DONNÉES » avec C. Duvette, P. Kervegan (INHA), W. Charles, S. Poublanc (Univ. Toulouse J. Jaurès) et C. Kermorvant (TEKLIA)

Mardi 7 novembre 2023 (9h30-12h00) en salle 70 – Buffet repas au Belvédère (12h15-13h15)

9h30-12h00. **SESSION** « DONNÉES HISTORIQUES NUMÉRIQUES. ÉVALUATIONS CRITIQUES ET PERSPECTIVES D'UTILISATIONS » avec L. Kesztenbaum (INED), G. Postel-Vinay (PSE), C. Lemerancier (CNRS).

Mardi 7 novembre 2023 (13h30-17h30) au BnF DataLab

13h30-17h30. **ATELIERS-FORMATIONS** SoDUCo – BnF DATA LAB (sur inscription, places limitées)

DONNÉES HISTORIQUES NUMÉRIQUES

ÉVALUATIONS CRITIQUES ET PERSPECTIVES D'UTILISATIONS

Mardi 7 novembre 2023 (09h30-12h00 en salle 70).

La matinée « Données historiques numériques » vise à confronter l'expérience de chercheurs qui construisent et analysent régulièrement des corpus de données numériques élaborés à partir de documents d'archives (texte et cartes). Dans un moment où l'on assiste à l'explosion de la production de données numériques ouvertes issues de sources anciennes, les questions classiques des conditions d'appropriation et de réutilisation de ces nouvelles « sources » par des tiers se posent de manière accrue.

Cette matinée invite plus particulièrement à réfléchir à ces questions dans le cas où les corpus de gros volumes de données numériques sont produits à partir de processus entièrement automatisés d'extraction de documents et d'enrichissement. Partant des travaux réalisés dans le cadre du programme SoDUCo, notamment du corpus des *Annuaire historiques parisiens* et de la *géolocalisation historique* des entrées d'annuaires sur les plans anciens, nous souhaitons aborder les questions suivantes au cours de la matinée : Comment et à quelles conditions peut-on s'emparer efficacement de telles « sources numériques » ? Comment les comprendre, évaluer leur « qualité » et leur rapport avec les sources initiales ? Comment comparer ou rendre cumulable les travaux très divers qui peuvent être issus de l'exploitation de ces jeux de données ? Enfin, nous souhaitons discuter collectivement les perspectives d'analyse et d'exploitation du corpus des *Annuaire historiques parisiens*.

Avec la participation de L. Kesztenbaum (INED), G. Postel-Vinay (PSE), C. Lemerancier (CNRS).

DONNÉES HISTORIQUES NUMÉRIQUES

ÉVALUATIONS CRITIQUES ET PERSPECTIVES D'UTILISATIONS

Mardi 7 novembre 2023 (09h30-12h00) en salle 70

09h30. **SoDUCo team**, *Les données géohistoriques produites au cours du programme SoDUCo. Evaluation d'une « source numérique ».*

10h00. **L. Kesztenbaum** (*Institut national d'études démographiques, INED*) et **G. Postel-Vinay** (*Paris School of Economics, PSE*), *Construction de bases de données historiques: retours d'expérience.*

10h30. Questions

10h45. Pause - café

11h00. **C. Lemerrier** (*CNRS, Centre de sociologie des organisations - Sciences Po-Paris*), *Histoire quantitative, données ouvertes en SHS.*

11h30. Questions et discussion finale

12h00. Déjeuner. Buffet repas au Belvédère (12h15-13h15)

4^E SÉANCE DE L'ATELIER SoDUCo-BnF

LES ATELIERS-FORMATION SoDUCo – BnF DATA LAB

Mardi 7 novembre 2023 (13h30-17h30).

BnF-DataLab - Bibliothèque nationale de France, Site François Mitterrand, Paris.

(Accès sur inscription ou sur invitation)

En complément de la 4^{ème} séance de l'Atelier SoDUCo-BnF, les membres du consortium SoDUCo organisent 3 ateliers-formation le mardi 7 novembre après-midi avec l'aide du BnF-DataLab. Ces ateliers-formations, accessibles sur invitation ou inscription, sont organisés en parallèle. Ils ont été imaginés pour permettre à différents publics de s'approprier et de réutiliser quelques-unes des ressources produites dans le cadre du programme de recherche SoDUCo. Chaque atelier proposera un scénario général d'utilisation et s'attachera ensuite à sa mise en pratique avec les stagiaires.

Avec la participation de M. Carlin, L.-A. Charles et de l'équipe du BnF DataLab

4^E SÉANCE DE L'ATELIER SoDUCo-BnF

LES ATELIERS-FORMATION SoDUCo – BnF DATA LAB

Mardi 7 novembre 2023 (13h30-17h30). Accueil à partir de 13h15 dans le Hall Est de la BnF

BnF-DataLab - Bibliothèque nationale de France, Site François Mitterrand, Paris.

13h30. Début des ateliers au DataLab :

ATELIER N°1 – *Exploration des données et cartographie grand public*

Programme : Comment accéder aux données et travailler avec les ressources SoDUCo. Org. P.-A. Le Ny & M. Hersent & J. Perret & P. Cristofoli - Salle 12 pers.

ATELIER N°2 – *Utiliser la chaîne SoDUCo d'extraction et d'enrichissement de documents textuels numériques.* Org. : E. Carlinet & J. Chazalon & B. Dumenieu, [& N. Abadie] - Salle 6 pers.

ATELIER N°3 – *Graphes de Connaissance Géohistoriques Professionnels.* Org. N. Abadie & S. Tual & J. Gravier - Salle 6 pers.

16h30. Pause

16h45. Bilan collectif et discussion des ateliers (Mezzanine du DataLab)

17h30. Clôture de la journée.

Annuaire historiques parisiens, 1798-1914.

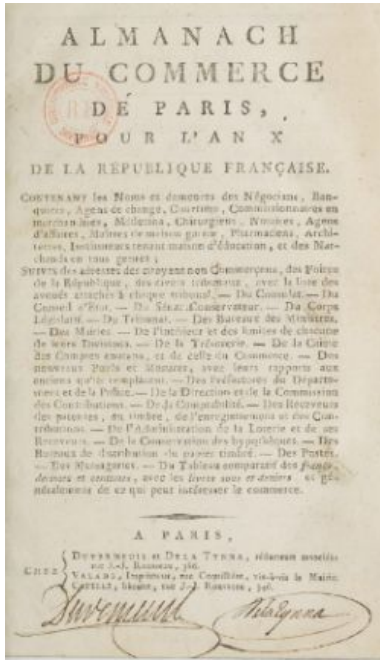


Conditions d'appropriation et de réutilisation

Annuaire historiques parisiens, 1798-1914. Extraction structurée et géolocalisée à l'adresse des listes nominatives [par ordre alphabétique et par activité] dans les volumes numérisés, SoDUCo Team, Version III - Septembre 2023, Paris, DataSet Nakala, <https://nakala.fr/10.34847/nkl.98eem49t>

Ouvrages, séries et collections d'annuaires commerciaux (XIX^e)

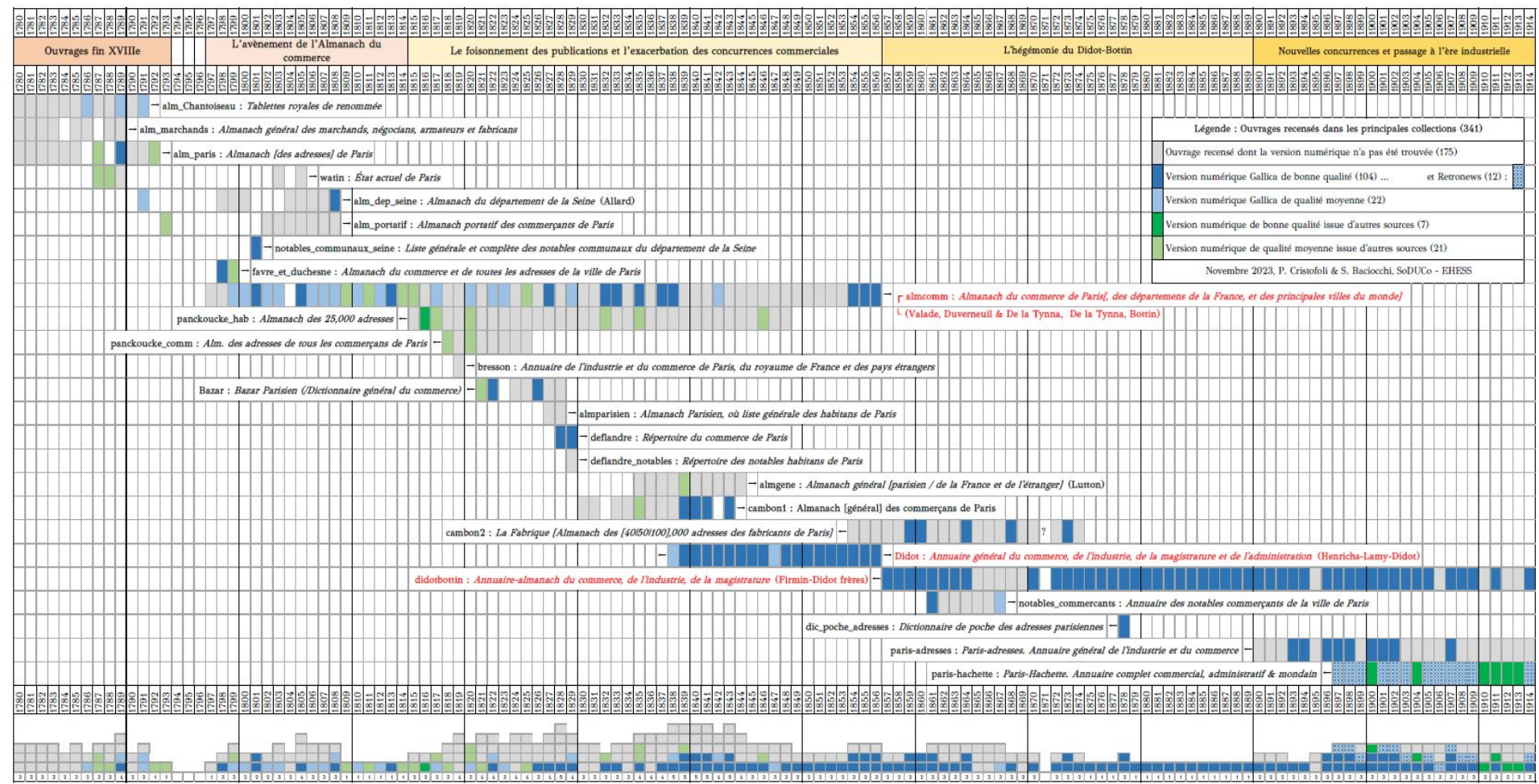
- Définition d'un corpus des *Almanach et Annuaires commerciaux parisiens du XIX^e*: Recensement des « ouvrages proposant des listes d'adresses d'individus ou d'établissement, de leurs états et/ou activités professionnelles respectives et localisés dans la ville de Paris entre 1780 et 1914 ».
- Recensement des listes d'adresses contenues dans ce corpus.



	Les collections d'ouvrages proposant des listes d'adresses parisiennes entre 1780 et 1914 (V. 2023-11, 25 collections recensées)	Code collection	Année min. obs.	Année max. obs.	Nb d'ouvrages recensés	Nb. Ouvr. avec version numérique	Nb. Ouvr. sans version numérique	% Numérisation collection	Gallica bonne qualité	Bertrons bonne qualité	Gallica qualité moyenne	Autre source bonne qualité	Autre source qualité moyenne
1	<i>Tablettes royales de renommée</i>	alm_chantoiseau	1780	1791	12	3	9	25%				3	
2	<i>Almanach général des marchands, négocians, armateurs et fabricans</i>	alm_marchands	1780	1789	8	0	8	0%					
3	<i>Almanach [des adresses] de Paris</i>	alm_paris	1780	1792	13	3	10	23%	1				2
4	<i>État actuel de Paris</i>	watin	1787	1805	5	2	3	40%					2
5	<i>Almanach du département de la Seine</i>	alm_dep_seine	1791	1808	9	2	7	22%	1		1		
6	<i>Almanach portatif des commerçans de Paris</i>	alm_portatif	1793	1808	8	1	7	13%					1
7	<i>Liste générale et complète des notables communaux du département de la Seine</i>	notables_communaux_seine	1801	1801	1	1	1	100%	1				
8	<i>Almanach du commerce et de toutes les adresses de la ville de Paris</i>	favre_et_duchesne	1798	1799	2	2	100%	1					1
9	<i>Almanach du commerce de Paris, des départemens de la France, et des principales villes du monde</i>	almcomm	1797	1855	59	32	27	54%	12		15		5
10	<i>Almanach des 25,000 adresses des principaux habitans de Paris</i>	panckoucke_hab	1815	1848	34	6	28	18%				1	5
11	<i>Almanach des adresses de tous les commerçans de Paris</i>	panckoucke_comm	1818	1825	8	2	6	25%					2
12	<i>Annuaire de l'industrie et du commerce de Paris, du royaume de France et des pays étrangers</i>	bresson	1819	1819	1	0	1	0%					
13	<i>Bazar Parisien (-> Dictionnaire général du commerce)</i>	bazar	1821	1828	7	3	4	43%	2				1
14	<i>Almanach Parisien, ou liste générale des habitans de Paris</i>	almparisien	1827	1828	2	0	2	0%					
15	<i>Répertoire du commerce de Paris</i>	deflandre	1828	1829	2	2	0	100%	2				
16	<i>Répertoire des notables habitans de Paris</i>	deflandre_notables	1829	1829	1	0	1	0%					
17	<i>Almanach général [parisien / de la France et de l'étranger]</i>	almgene	1835	1844	10	1	9	10%					1
18	<i>Almanach [général] des commerçans de Paris</i>	cambon1	1830	1843	12	5	7	42%	4				1
19	<i>La Fabrique puis Almanach des 4050/100,000 adresses des fabricants de Paris</i>	cambon2	1854	1874	20	5	15	25%	5				
20	<i>Annuaire général du commerce, de l'industrie, de la magistrature et de l'administration</i>	didot	1838	1856	19	19	100%	17			2		
21	<i>Annuaire-almanach du commerce, de l'industrie, de la magistrature</i>	didotbottin	1857	1914	57	47	10	82%	47				
22	<i>Annuaire des notables commerçans de la ville de Paris</i>	notables_commerçants	1861	1867	7	2	5	29%	1		1		
23	<i>Dictionnaire de poche des adresses parisiennes</i>	dic_poche_adresses	1878	1878	1	1	100%	1					
24	<i>Paris-Adresses</i>	paris_adresses	1890	1914	25	9	16	36%	9				
25	<i>Paris-Hachette</i>	paris_hachette	1897	1914	18	18	0	100%		12			6
	Ensemble		1780	1914	341	166	175	49%	104	12	22	7	21

cf. *Inventaire critique du corpus des collections d'ouvrages proposant des listes d'adresses parisiennes au XIX^e siècle*, S. Baciocchi & P. Cristofoli (CRH LaDéHiS - EHES). (Actualisation du document présenté lors de la 2e journée de l'atelier SoDUCo-BnF, 10 novembre 2022)

INVENTAIRE CHRONOLOGIQUE DES GRANDES COLLECTIONS :



STRUCTURATION DES INFORMATIONS

L'information diffusée/recherchée : « triplets » "NOM" – "ACTIVITÉ" – "ADRESSE"

L'information éditoriale : « entrées d'annuaires », blocs d'entrées, pages, plages de pages, listes d'adresses.

ListNoms

"nom" – "activité" – "adresse"

ListProfs

"nom" – ["activité"] – "adresse"
au sein de "catégories d'activités"

ListRues

"nom" et "activité" situé(s) à un
"numéro" au sein d'une catégorie "rue"

LISTE GÉNÉRALE

ADRESSES DE PARIS,

ET DES PRINCIPAUX ÉTABLISSEMENTS DE CETTE CAPITALE.

Abelin, architecte, Marivaux-Temple, 60. | Acard, libraire, Casette, 32.
Audié fils, passementier, Bouclier, 6 (R). | Acaud et Gallego, pharmacien, Puits, 19.

Didion (C.), O. *, inspecteur général des ponts et chaussées, directeur du chemin de fer d'Orléans, St-Lazare, 101.

Didot (N.C.), bois à ouvrer, boulev. Mazas, 5.
Didot (Mme), brunisseuse, passage de la Marmite, 14, escalier F.

Didot (Mme), couturière, Ecole-de-Médecine, 101.

Didot (Henri), gantier, Nve-St-Denis, 16.

Didot (F.), gantier, Faub.-St-Honoré, 56.

Didot, lettres en relief, pass. Verdeau, 3.

Didot, tabac, Dragon, 22.

Didot (Ch.), vins en gros et consignat., St-Christophe, 8.

Didot (Ambr.-Firmin), O. *, N.C., membre du conseil municipal de Paris, Jacob, 56.

Didot (Alfred-Firmin), Varenne, 57.

Didot (Hyacinthe-Firmin) *, quai Voltaire, 21.

Didot (Paul-Firmin), Bouaparte, 1.

Didot (Mme Fréd.), quai Malaquais, 3.

Didot (P. H.), anc. banquier, Cherche-Midi, 11.

Didot (Firmin), O. *, frères (N.C.), fils et Cie, imprim. en lettres et libraires, Jacob, 56; de

Adam (H.), peintre, Malgoune, 3.
Adam, peint. vitr., St-Germain-des-Prés, 3.
Adam, pâtisier et maison, Descartes, 24.
Adam (H. E.), impr., Nve-Saint-Pierre-Ch., 6.
Adam, serrurier, quai des Augustins, 55.
Adam, sous-chef à l'épicerie, Grand-Casnet, 9.
Adam, tabac, Noyers, 19.
Adam, tourneur sur métaux, Henry, 2.
Adam, vins, Provençaux, 22.
Adam et Lempic, bijoutier, Quincampoix, 64 (R).
Adam-Zeller, serrur. en vitr., Cadet, 10.
Adami, peintre-vitr., Faub.-Stain-Denis, 100 (R).
Adin, journaliste-douanier, pl. de la Madeleine, 6.
Aldie, libraire, boulev. Poissonnière, 17.
Aldie-Margou, méd., Faub.-Poissonnière, 4.
Aldie, prof. au collège Henri IV, Capoue, 34.
Aldinger (R. F.), impr., Bretagne, 6.
Aldie jeune, maçon-plâtrier, Jour, 4.
Aldie, rev. – Vieux, 4.

PROFESSIONS

DES

COMMERCANTS ET INDUSTRIELS DE LA VILLE DE PARIS.

PAR ORDRE ALPHABÉTIQUE.

ABAT-JOUR. — ACHAT. — ACIER POLI.

CARROSSIERS. Voy. Constructeurs de voitures.

CARTES GÉOGRAPHIQUES ET CÉLESTES, GLOBES ET SPHÈRES.

Ardriveau-Goujon (E.) (N.C.), grand assortiment de cartes françaises et étrangères, globes et sphères, cartes du Dépôt de la guerre, etc., Bac 21.

Barthélemy (J.), quai des Grands-Augustins, 39.

Baerckeller et Cie (Hooek-Baerckeller, successeur), cartes géographiques en relief, méd. 2e cl. 1855, Vendôme 24.

CARTES A JOUER (FABR.).

Arnoult (Vve) et fils, Fossés-St-Victor, 24. Avril et Cie, Croix-des-Petits-Champs, 37, fab. Grande-Rue, 47, à Vaugirard. Chevalier et Cie, Renard-St-Sauveur, 3.

De La Rue (Th.) et Cie (de Londres), rue du Château-d'Eau, 52 (Voir la Revue industrielle.) Delèdre fils, successeur de son père, cartes à jouer, cartes en feuilles et collage à façon, rue de Bèver, 31, au coin du boulev. St-Germain.

Grimaud (B.-P.), Bondy, 70. Julieonne (Mme) Nve-St-Merri, 23. Malmenayde (E.), cartes à jouer.

Didiot (Ch.), vins en gros et consignat., St-Christophe, 8.
Didot (Ambr.-Firmin), O. *, N.C., membre du conseil municipal de Paris, Jacob, 56.
Didot (Alfred-Firmin), Varenne, 57.
Didot (Hyacinthe-Firmin) *, quai Voltaire, 21.

LISTE GÉNÉRALE DES ADRESSES DE PARIS

CLASSÉES PAR ORDRE ALPHABÉTIQUE DE RUES

ET PAR NUMÉROS DE MAISON.

AGACIAS (petite rue des),
5 Lecat (A.), charp.

AGUESSEAU (l').

- Sentinelle de l'Armée (la) (journal).
- Mauduit (de), rédacteur en chef.
- Jacquinet, chaudronnier. Herbet, teint.-dégr.
- Cotereau (Mme), prof. de musique.
- Cintrat, O. *, sous-dir. aux affaires étrang.
- Coutelet, serrurier.
- Deffaux, tonnelier.
- Sevestre, fabr. pain-d'é.
- Petitjean *, référ. de la cour des comptes.
- Basset *, rentier.
- Levique-Lascombe, ti-buairie anglaise.
- Beauvy (Baron de). Corneville (Comte de).
- Quirieu (Mise de).
- Cowens, taverne angl.

- 15 Espagnac (Cte d').
- 16 Danneville jeune, port-celaire et faïence.
- 18* Aigle (Marquis de l'). Aigle (Arthur de l'). Aigle (Victor de l'). Moges (Comte de), C. *, contre-amiral. Moges (Comtesse de).
- 20 Girardin (comte de) *, lieutenant général.
- 22 Esnoult (St-Sernin). Lanarre (Cte de).

AIQUILLERIE (de l').

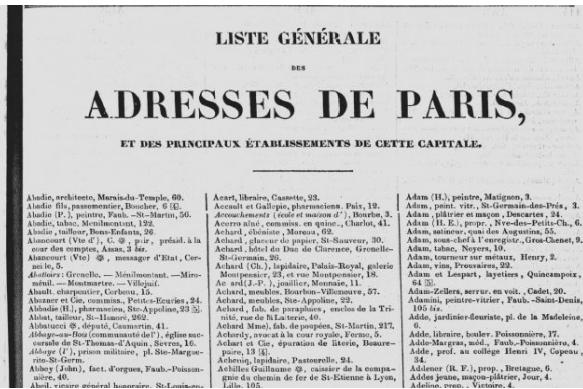
- 5* Courvaux, coiffeur. Collin (Mme), lingerie. Bourot (L.), boucher. Lemarchand (J.-H.), médecin-accoucheur. Lionnet, architecte. Jean, aubergiste.
- 6 Leblanc – Prudhomme (V.), passementerie. Audouin (L.), papeterie en gr. Girard (Vve), autoumm. Loir fils, épice et comm.

EXTRACTION SoDUCo DES INFORMATIONS

ListNoms



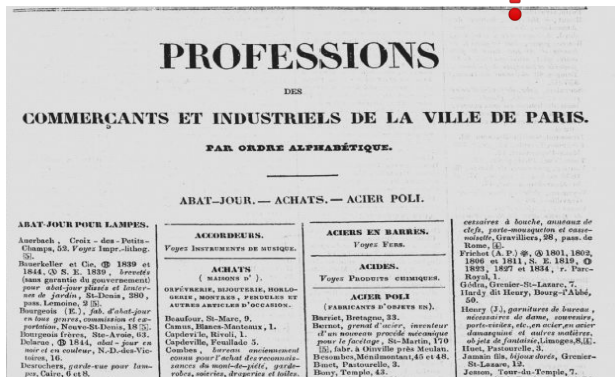
"nom" - "activité" - "adresse"



ListProfs



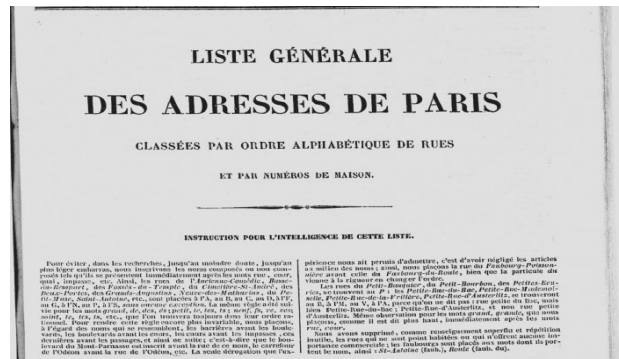
"nom" - ["activité"] - "adresse"
au sein de "catégories d'activités"



ListRues



"nom" et ["activité"] situé(s) à un
"numéro" au sein d'une catégorie "rue"



Traitement systématique

Intérêt : listes simples et suites éditoriales et chronologiques.

Attention: apparition tardive des premières listes "générales" d'adresses de Paris: Coll. Panckroucke_hab (1815), Coll. Almcomm (en 1824)

Traitement systématique :

En partie exploitables :

- Actuellement, pas de post-traitement ré-associant les catégories d'activités aux entrées qui en dépendent (en cours).
- Accès par sélection de plages de pages possible

Non traitées.

Structure d'information non prise en compte pour l'instant par l'extraction.

Cf. autres projets : Annuaires de propriétaires et des propriétés de Paris (1898),

Les données "Annuaire"

Annuaire historiques parisiens, 1798-1914.

Version 3, Septembre 2023

(22 741 928 entrées d'annuaire)

<https://nakala.fr/10.34847/nkl.98eem49t>

Quelques "Partis pris" du programme de recherche SoDUCo

- préserver le lien aux sources (à l'échelle de l'entrée et de la page)
- garder la trace des traitements (chaîne de traitement des sources et d'enrichissement des données produites)
- proposer une spatialisation historique des données
- méthode progressive (versions de données successives avec évaluation et feed-back pour ajuster la chaîne de traitements).

V2, juin 2022 : 9 825 018 entrées / ListeNoms : 42 128 pages traitées, 7 494 962 entrées / ListProfs : 17 707 pages traitées, 2 330 056 entrées

Annuaire recensés	341
Annuaire recensés avec version numérique	166
Annuaire traités	144
Pages traitées	143 354

Listes alphabétiques

Listes recensées *	135
Listes traitées	132
Pages traitées	54 142
Entrées	10 253 910

Listes professionnelles

Listes recensées *	135
Listes traitées	134
Pages traitées	89 212
Entrées	12 490 018

Listes d'adresses

Listes recensées *	79
--------------------	----

* sans distinction des diverses listes alphabétiques, professionnelles ou d'adresses pouvant exister dans un même Annuaire

Différentes voies d'accès aux données numériques

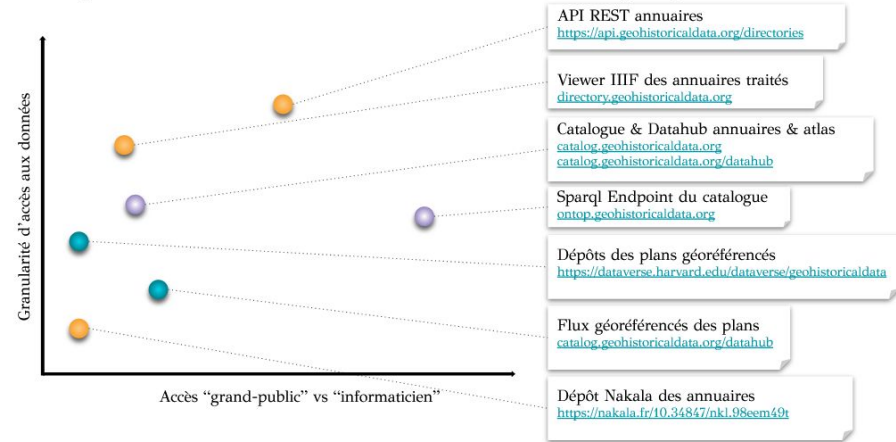
Taille du corpus et des ensembles de données produites.

- Des données fines
- ... et en masse
- Des données longitudinales inédites ...

Un ensemble de contraintes pour la manipulation et l'analyse.

- Difficile de l'envisager / manipuler dans sa totalité.
- Accès spécifiques en fonction des questions et des problématiques (profils d'utilisation)
- Difficile d'évaluer la "qualité" des données

Ecosystème de diffusion des données du projet SoDUCo



Chaine de création des données : Fichiers Json et base PostgreSQL

Formats de sortie différents : Json, Géopackage, Csv

Entrepôt Nakala (versionnement):

- Fichiers par type de liste, Base globale (Gpkg)
- Fichiers par type de liste, annuelle (Gpkg, csv)

Viewer IIIF : accès par unité documentaire et export (csv json)

API REST : Interrogation directe par requête web (csv, json)

Évaluation critique d'une source numérique



les allers-retours

COMPARAISON INFORMATIONS DE LA SOURCE ET RESULTATS EXTRACTIONS

Didot_1847, Vues 21-22 : INTRODUCTION A L'ANNUAIRE GÉNÉRAL DU COMMERCE.

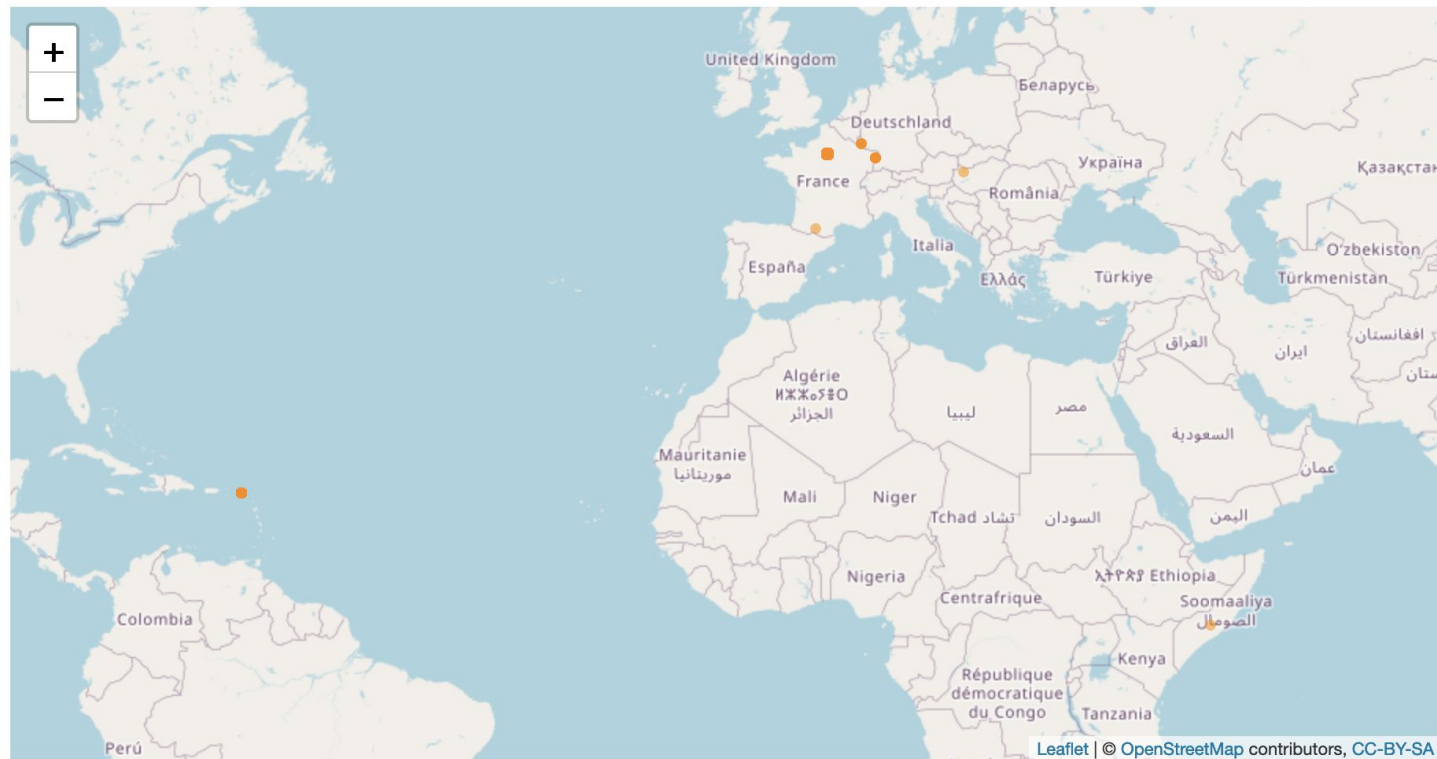
Le succès toujours croissant de l'ANNUAIRE GÉNÉRAL DU COMMERCE nous dispense de faire ressortir l'importance de cette vaste publication, qui s'augmente chaque année d'un certain nombre de feuilles, et dont le destin est de suivre pas à pas les progrès incessants de la fabrication française, et notamment de cette admirable industrie parisienne, sans rivalité dans le monde entier. [...]

L'accroissement du nombre des adresses qui figurent à la liste générale par ordre alphabétique est proportionnellement plus grand encore: en 1838, le nombre des adresses des habitants de Paris insérées dans l'Annuaire du Commerce s'élevait à 56,511 (occupant 207 pag.); ce nombre était de 72,864 (occupant 264 pages; en 1840); il est de 83,142 (occupant 298 p.) en 1847. [...]

Indexation ouvrages et listes								SoDUCo Traitements		Didot 1847 informations	
liste_type	Annee	code_ouvrage	collection	nbcol	npage_pdf_d	npage_pdf_f	nbvuespdf	soduco_pages	soduco_entrees	Didot1847_pages	Didot_1847_entrées
ListNoms	1838	Henrichs_1838	Didot	3	222	428	207	207	43 459	207	56 511
ListNoms	1840	Lamy_1840	Didot	3	128	417	290	290	65 123	264	72 864
ListNoms	1847	Didot_1847	Didot	3	321	618	298	298	65 301	298	83 142
ListProfs	1838	Henrichs_1838	Didot	4	86	221	136			140	
ListProfs	1840	Lamy_1840	Didot	4	417	614	198			190	
ListProfs	1846	Didot_1846	Didot	4	570	912	343			343	

Processus d'extraction & exploration des données

Géocodage : exemples de Saint-Martin et du Luxembourg

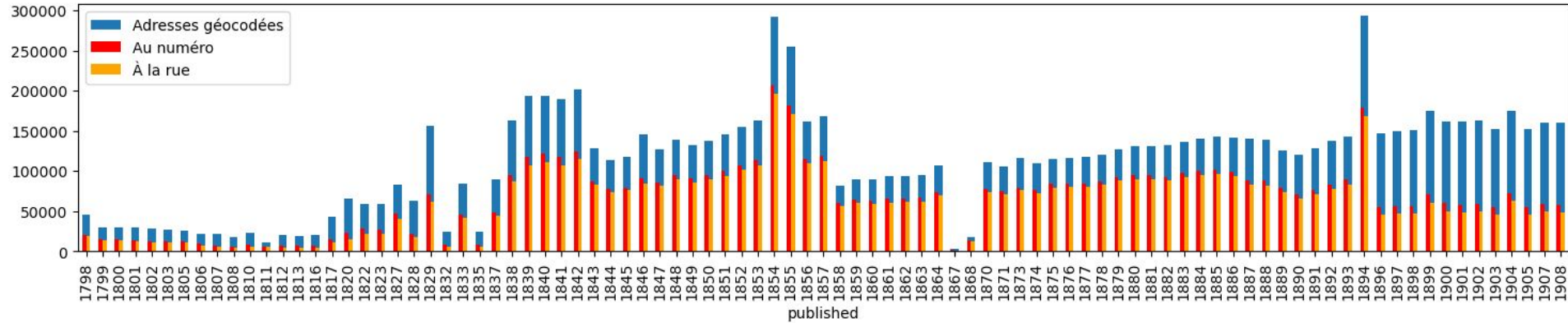


Liste alphabétique de l'annuaire de Lamy 1839, données V2, 2022-06-20

Processus d'extraction & exploration des données

Géocodage : V2, juin 2022

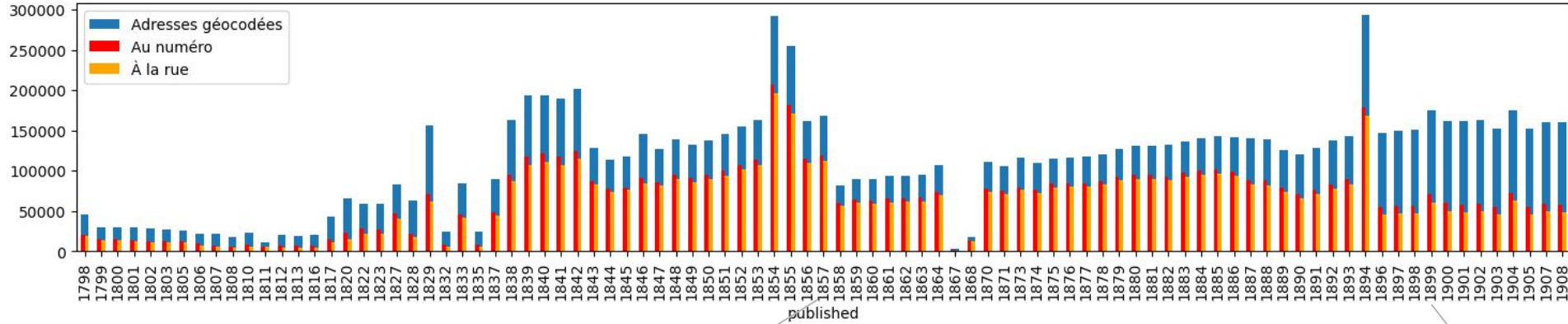
Adresses extraites par années de publications des annuaires et proportion géocodée, données V2, 2022-06-20



Processus d'extraction & exploration des données

Géocodage : V2, juin 2022

Adresses extraites par années de publications des annuaires et proportion géocodée, données V2, 2022-06-20



Adeline (Mlle), couturière, Rivoli, 210.
Adeline, prop. , Victoire, 12.
Adeline, propriétaire, Joubert, 41.
Adelmann, garde des collections à l'école
imperiale des mines, Enfer, 30.
Adelon, avocat cour impér., Mont-Thabor, 9.
Adelon O. ✱, medecin, prof. à la Faculté de
médecine, Bac, 32.

Didot-Bottin, 1857

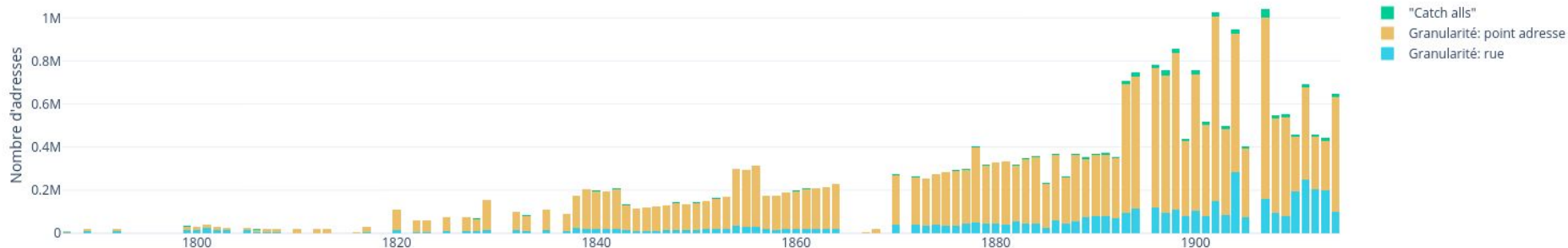
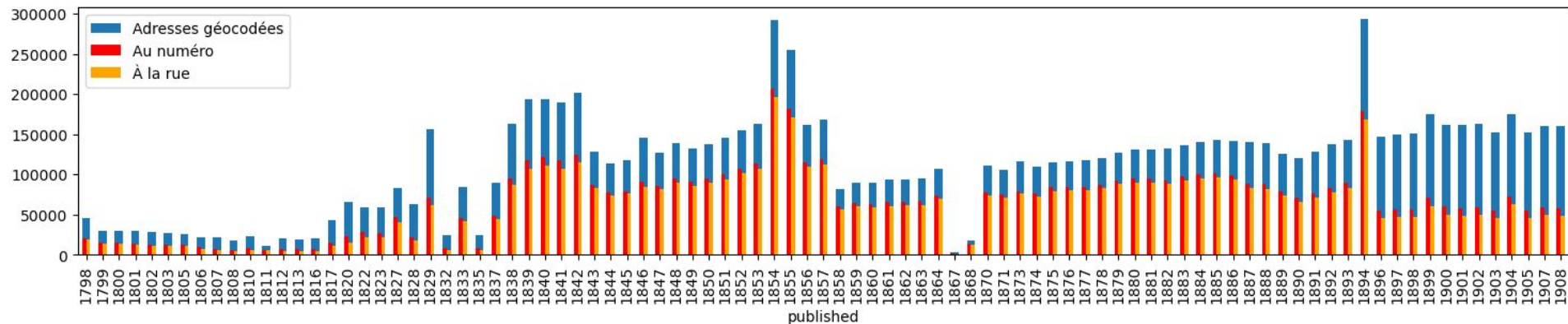
Adeline, vins, r. de Belleville, 251.
Adelis, épicier, r. Niepce, 2.
Adelmann (B.), boulanger, pass. St-Pierre-
Paul, 5.
Adelon (H.), av. Hoche, 2.
Adelon (M^{me}), r. de Courcelles, 24. **TÉLÉPH**
317.66.
Adelswärd (b^{ne} G. d'), boul. de Courcelles, 65.
TÉLÉPH 308.32!

Didot-Bottin, 1899

Processus d'extraction & exploration des données

Géocodage : V2, juin 2022 et V3, septembre 2023

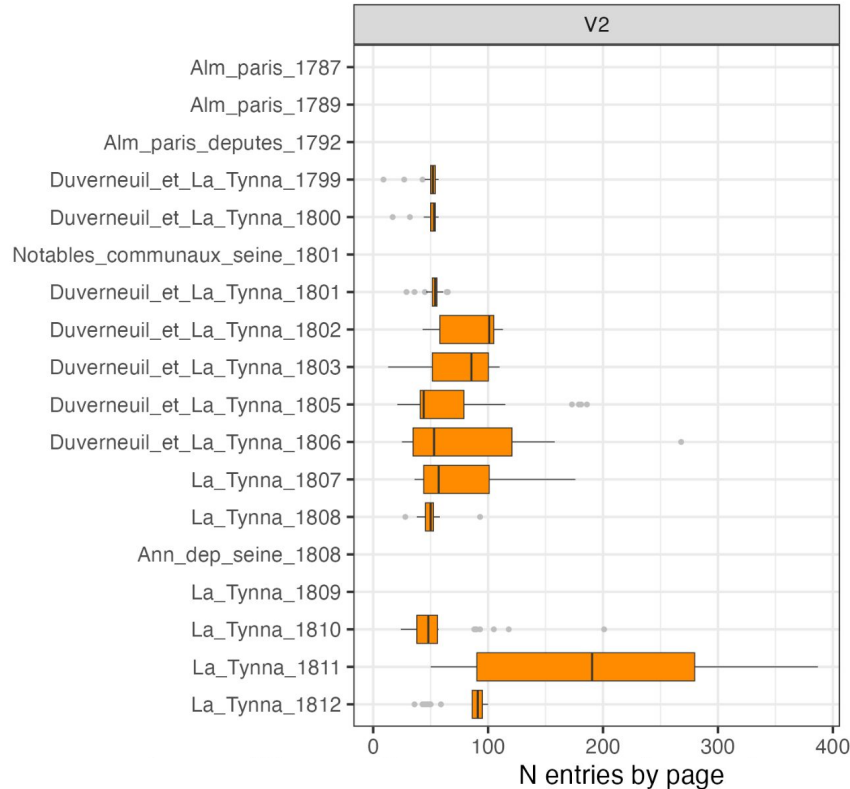
Adresses extraites par années de publications des annuaires et nombre d'adresses géocodées



Processus d'extraction & exploration des données

Segmentation des images et entrées multiples

Nombre d'entrées par page des listes alphabétiques



Processus d'extraction & exploration des données

Segmentation des images et entrées multiples

La Tynna, 1807

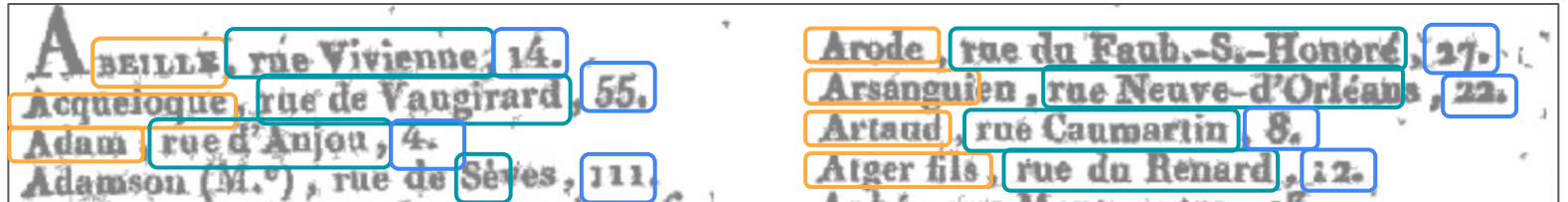
ABEILLE, rue Vivienne, 14.
Acqueioque, rue de Vangirard, 55.
Adam, rue d'Anjou, 4.
Adamson (M.^o), rue de Sèves, 111.

Arode, rue du Faub.-S.-Honoré, 27.
Arsanguien, rue Neuve-d'Orléans, 22.
Artaud, rue Caumartin, 8.
Atger fils, rue du Renard, 12.

V2, 2022-06-20

<PER> <LOC> <CARDINAL>

1 entrée



Processus d'extraction & exploration des données

Segmentation des images et entrées multiples

La Tynna, 1807

ABEILLE, rue Vivienne, 14.
Acqueoque, rue de Vaugirard, 55.
Adam, rue d'Anjou, 4.
Adamson (M.^o), rue de Sèves, 111.

Arode, rue du Faub.-S.-Honoré, 27.
Arsanguien, rue Neuve-d'Orléans, 22.
Artaud, rue Caumartin, 8.
Atger fils, rue du Renard, 12.

V2, 2022-06-20

<PER> <LOC> <CARDINAL>

1 entrée

The screenshot shows the text from the previous block with a single large blue bounding box around the entire text area. The text is: **A**BEILLE, rue Vivienne, 14. Acqueoque, rue de Vaugirard, 55. Adam, rue d'Anjou, 4. Adamson (M.^o), rue de Sèves, 111. Arode, rue du Faub.-S.-Honoré, 27. Arsanguien, rue Neuve-d'Orléans, 22. Artaud, rue Caumartin, 8. Atger fils, rue du Renard, 12.

V3, 2023-09-15

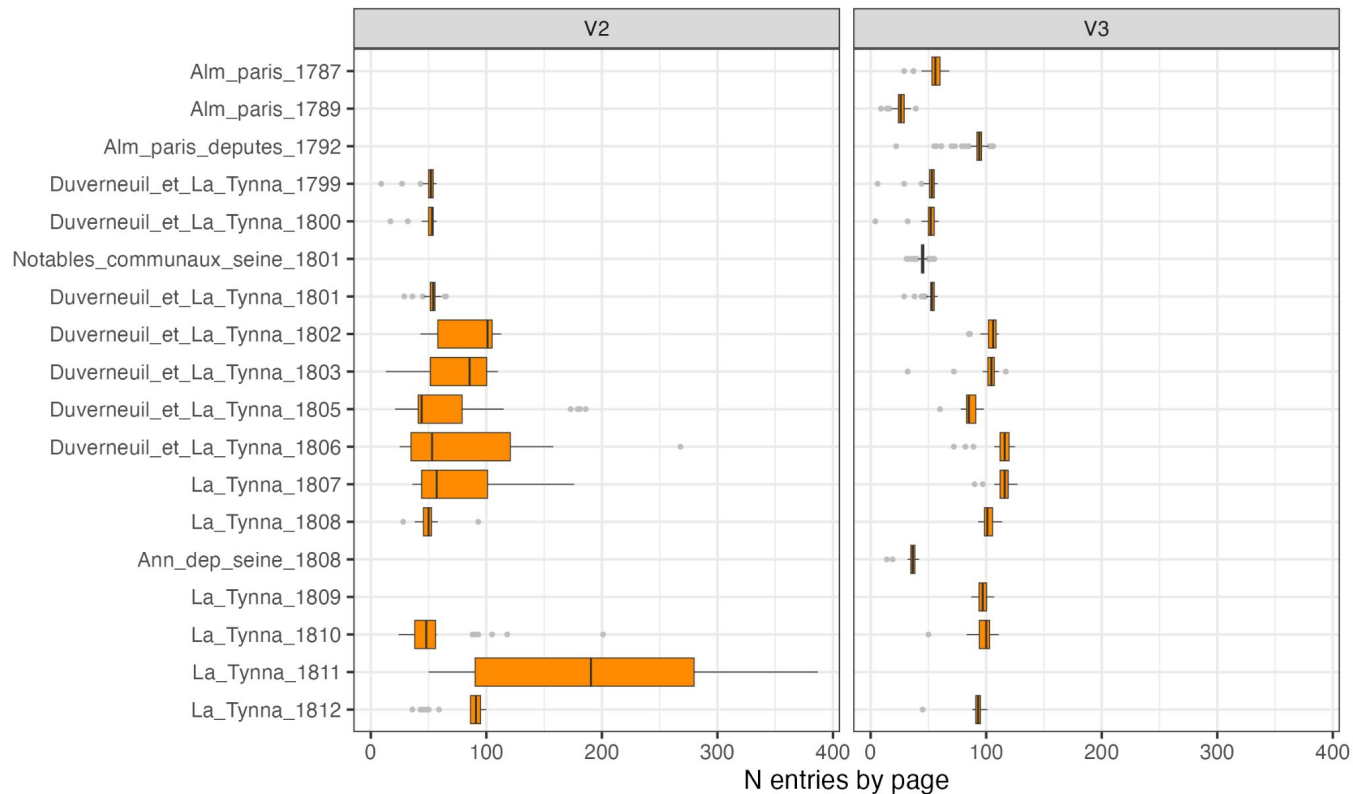
8 entrées

The screenshot shows the text from the previous block with 8 individual blue bounding boxes around each word and number. The text is: **A**BEILLE, rue Vivienne, 14. Acqueoque, rue de Vaugirard, 55. Adam, rue d'Anjou, 4. Adamson (M.^o), rue de Sèves, 111. Arode, rue du Faub.-S.-Honoré, 27. Arsanguien, rue Neuve-d'Orléans, 22. Artaud, rue Caumartin, 8. Atger fils, rue du Renard, 12.

Processus d'extraction & exploration des données

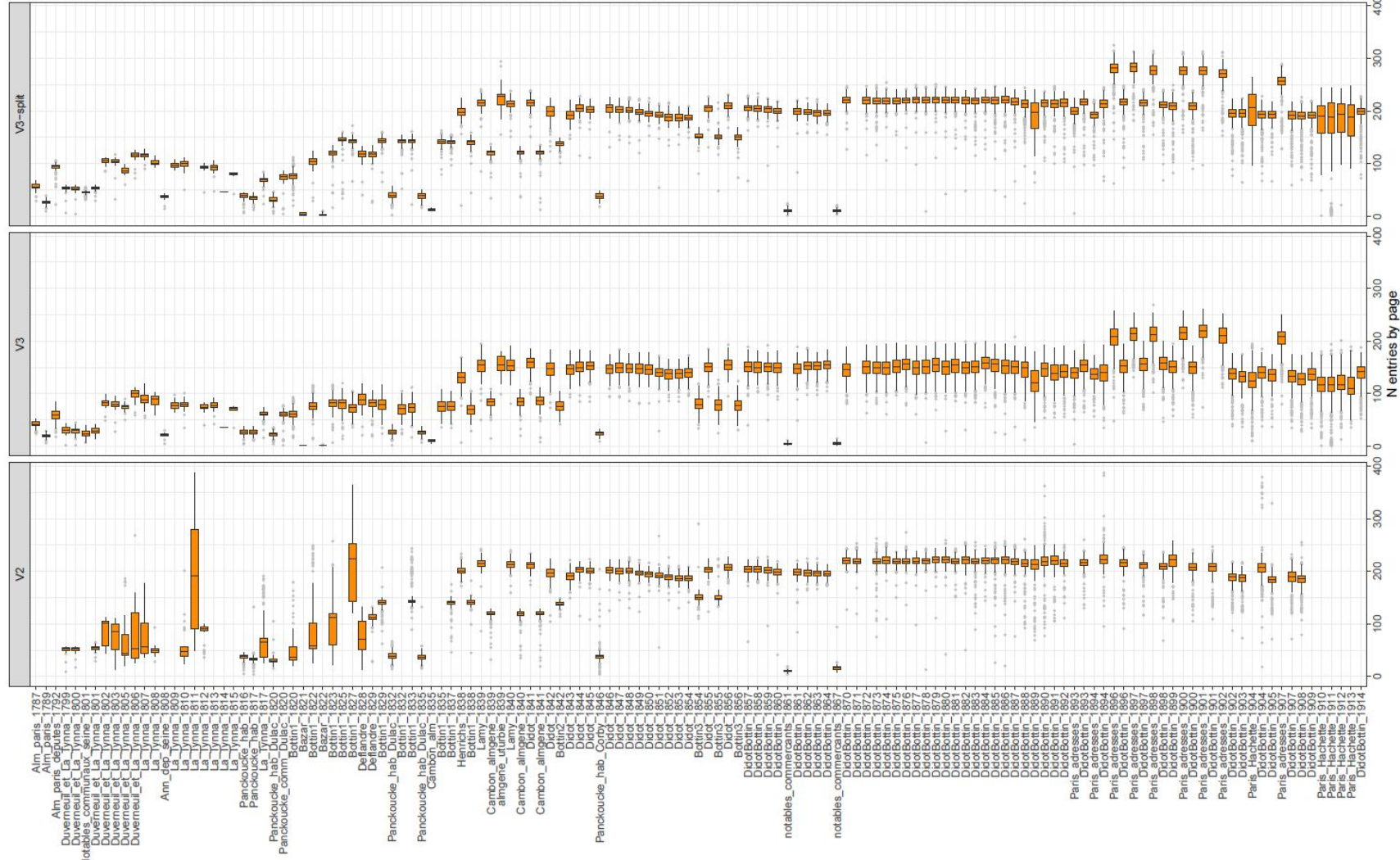
Segmentation des images et entrées multiples

Nombre d'entrées par page des listes alphabétiques et par version de données



Listes alphabétiques - Nombre d'entrées par page et par version de données (*)

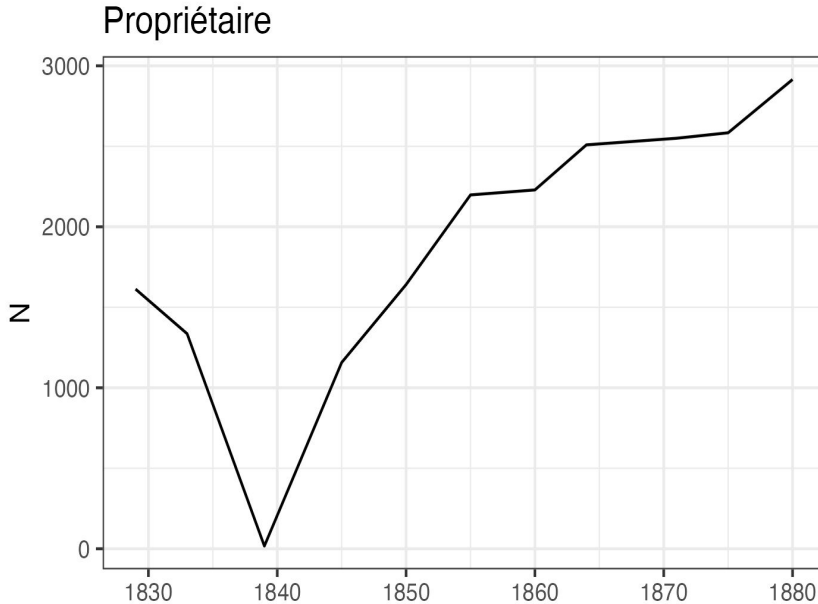
(V2 - Juin 2022, V3 - Sept. 2023, et V3 split (Sept. 2023, V3 avec post traitement))



Corpus de sources initiales & exploration des données

Autres
ouvrages

La disparition des propriétaires en 1840 ?
Effets de sources et effets de sélection



données V2, 2022-06-20, revues <https://doi.org/10.5281/zenodo.8388102>

Propriétaire Vs "éligible" :

Corpus d'étude (JG et MB)

Bottin 1833

Adam (F.-C.), *prop.*, Clichy, 50.

Panckoucke_hab_Dulac 1835, vue 51 -

Adam (F.-C.), *prop.*, *élig.*, r. de Clichy, 50.

Lamy 1839

Adam (F. - C.), *élig.*, Clichy, 50.

Cambon1_1839, vue 71 -

Adam, *prop.*, r. de Clichy, 50.

Alm_gene_uturbie_1839, vue 335 -

Adam (F.-É.), *propriétaire*, Clichy, 50.

Didot 1842, vue 191

Adam, *prop.*, Clichy, 50.

Bottin 1842, vue 19 -

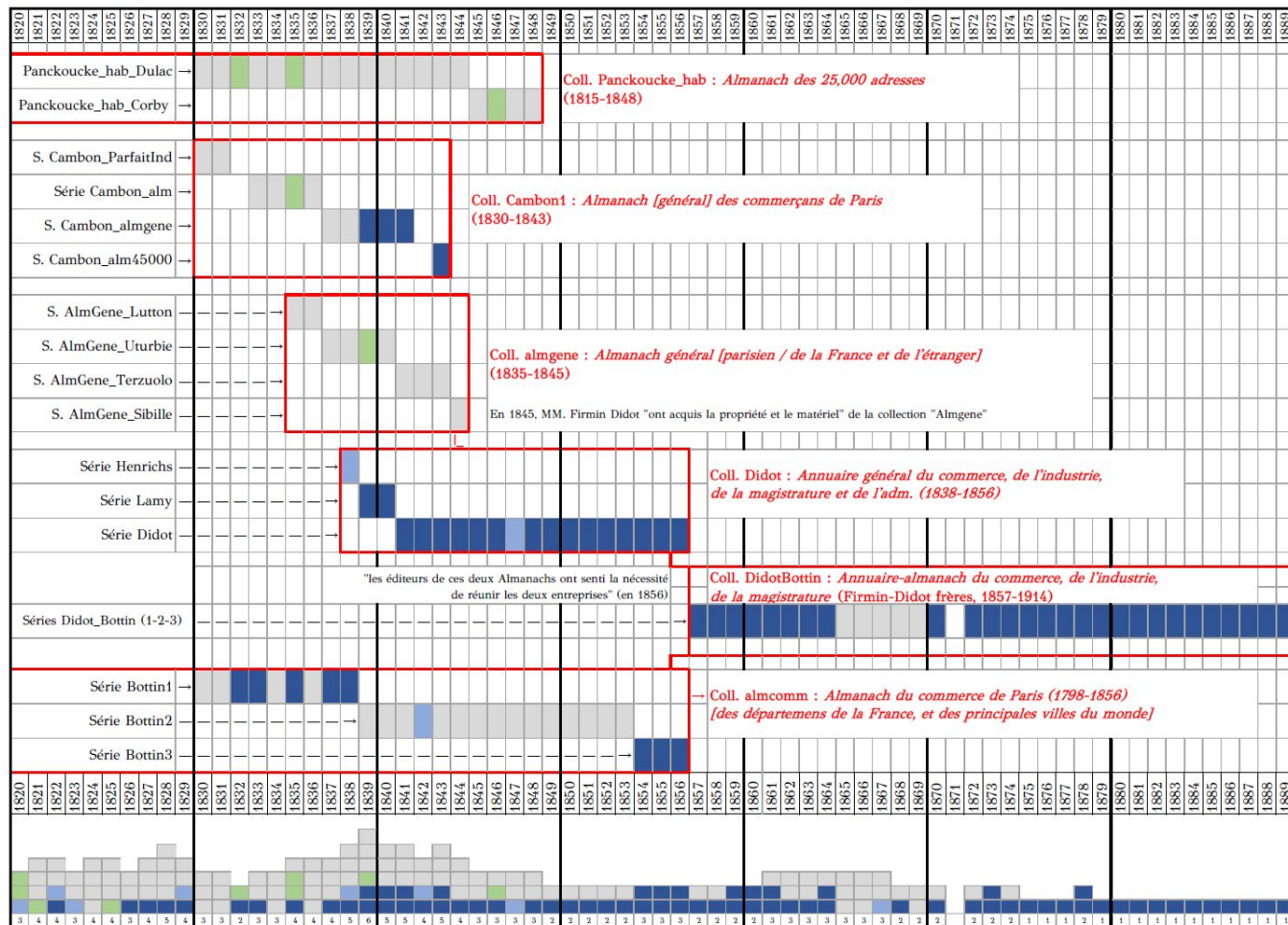
Adam (F.-C.), *prop.*, Clichy, 50.

Corpus de sources initiales & exploration des données

Des logiques éditoriales différentes et des sources d'information différentes pour les éditeurs d'annuaires. Henrichs/Lamy = nouvel annuaire dont la source principale est la liste électorale

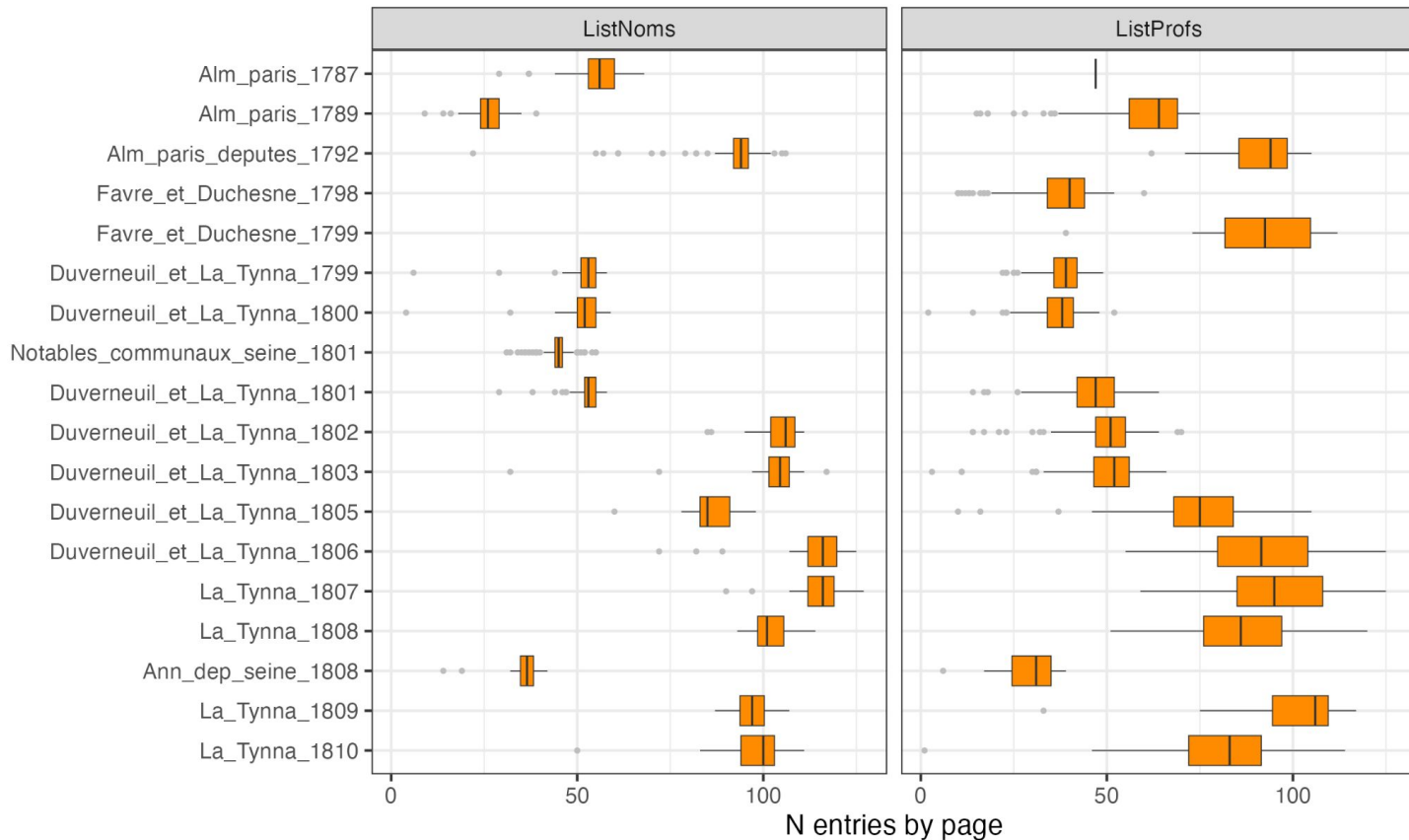
Cf. Henrichs, « Introduction », *Annuaire général du commerce, de l'industrie et de l'agriculture de France et des principales villes du monde*, 1838. Réf. : Henrichs_1838, [vue 12](#):

« [...] nous avons été puissamment secondés par M. le Ministre de l'intérieur, et par M. le préfet de la Seine, à qui nous devons notamment la communication des dernières listes électorales des départements. [...] »



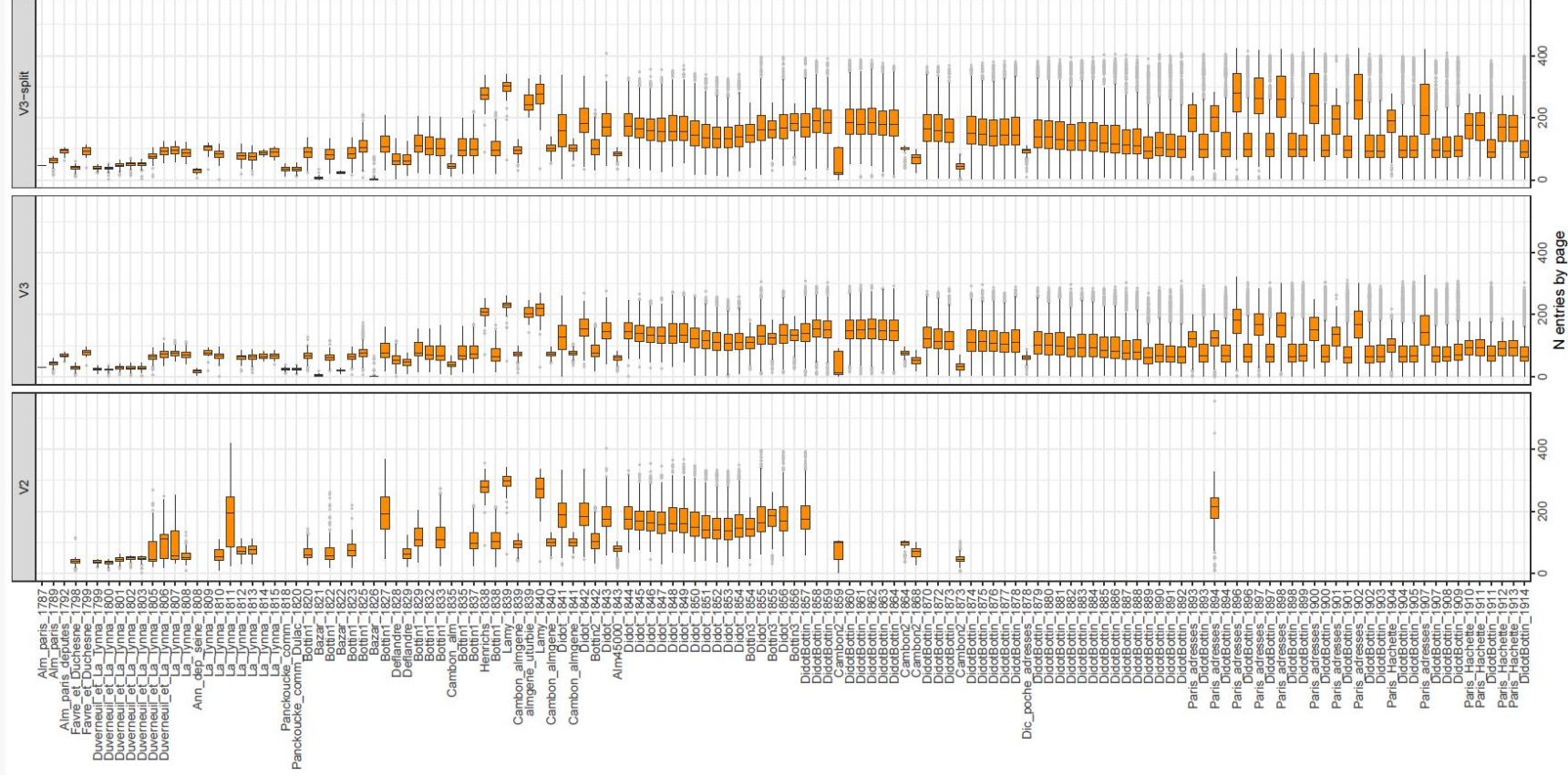
Corpus de sources initiales & exploration des données

Nombre d'entrées par page et par type de liste



Listes professionnelles - Nombre d'entrées par page et par version de données (*)

(V2 - Juin 2022, V3 - Sept. 2023, et V3 split (Sept. 2023, V3 avec post traitement))



Perspectives d'utilisations et d'analyses



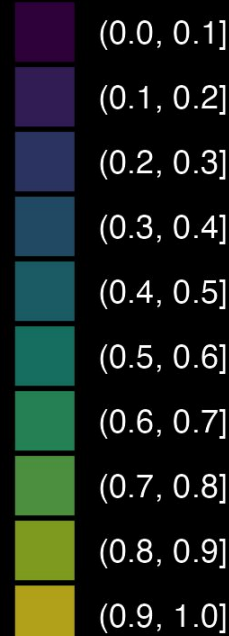
approches et échelles spatio-temporelles

Un *Annuaire* : une unité documentaire à un temps t

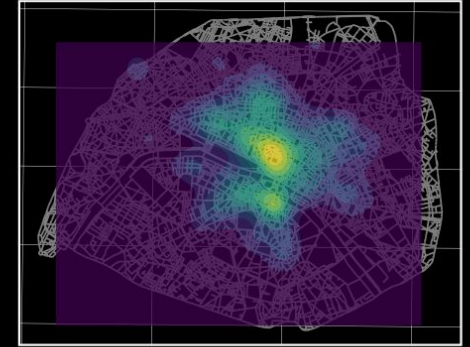
Didot 1850 (KDE)



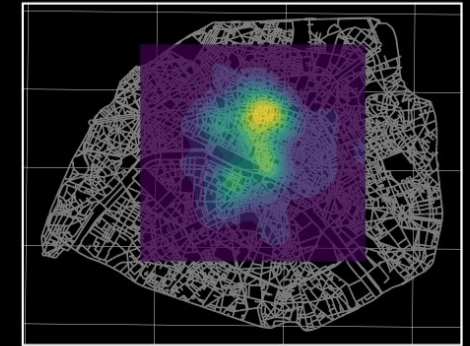
level



Epicier



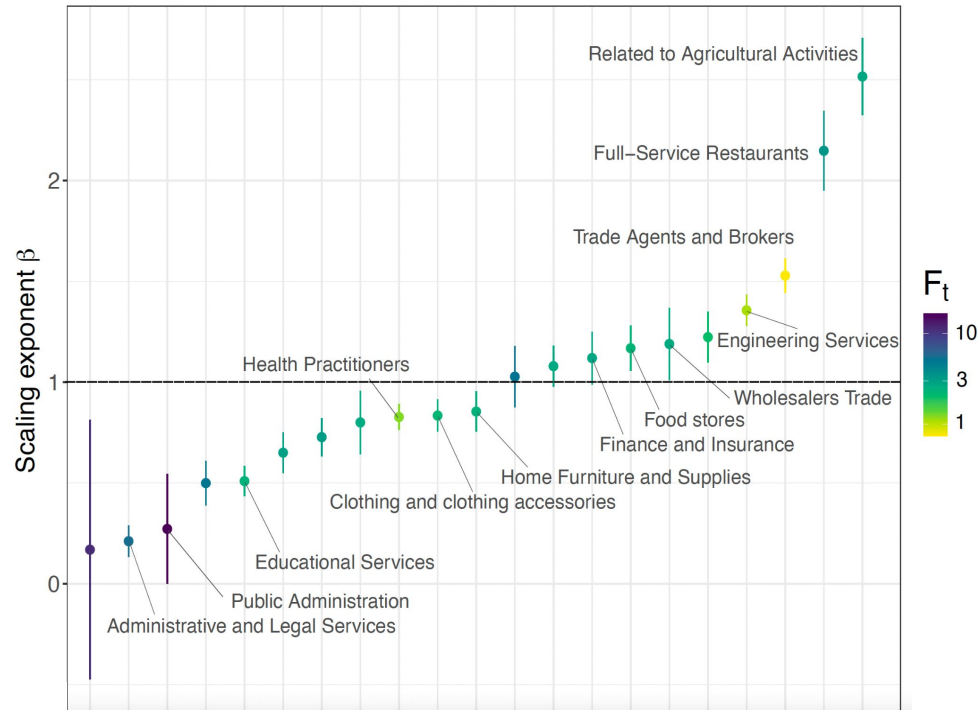
Rentier



Des *Annuaires* dans le temps : snapshots de la ville

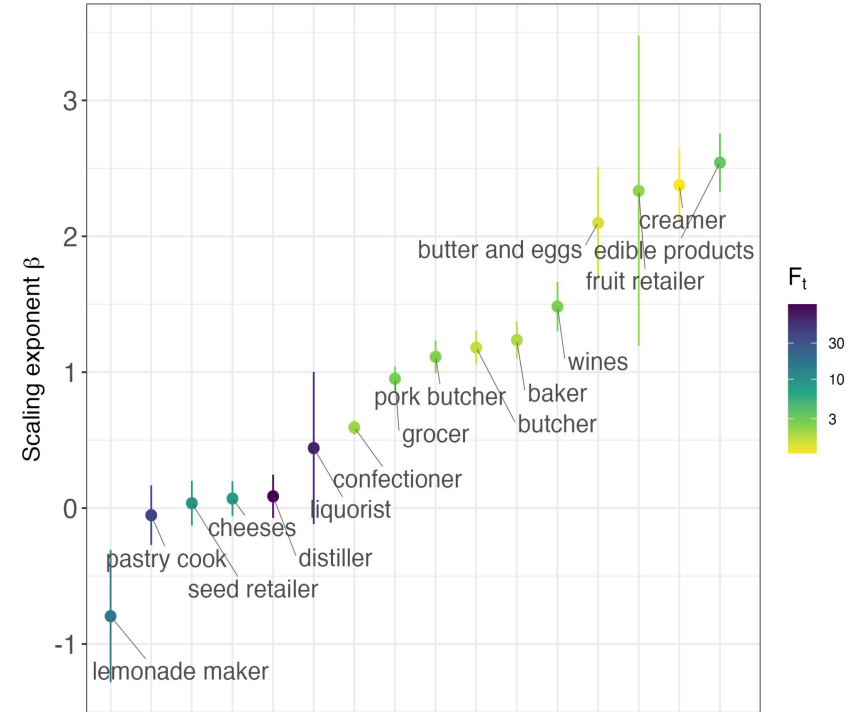
Activités et population de Paris entre 1833 et 1907, pas de temps ~ 5 ans

Reconstruction de catégories englobantes



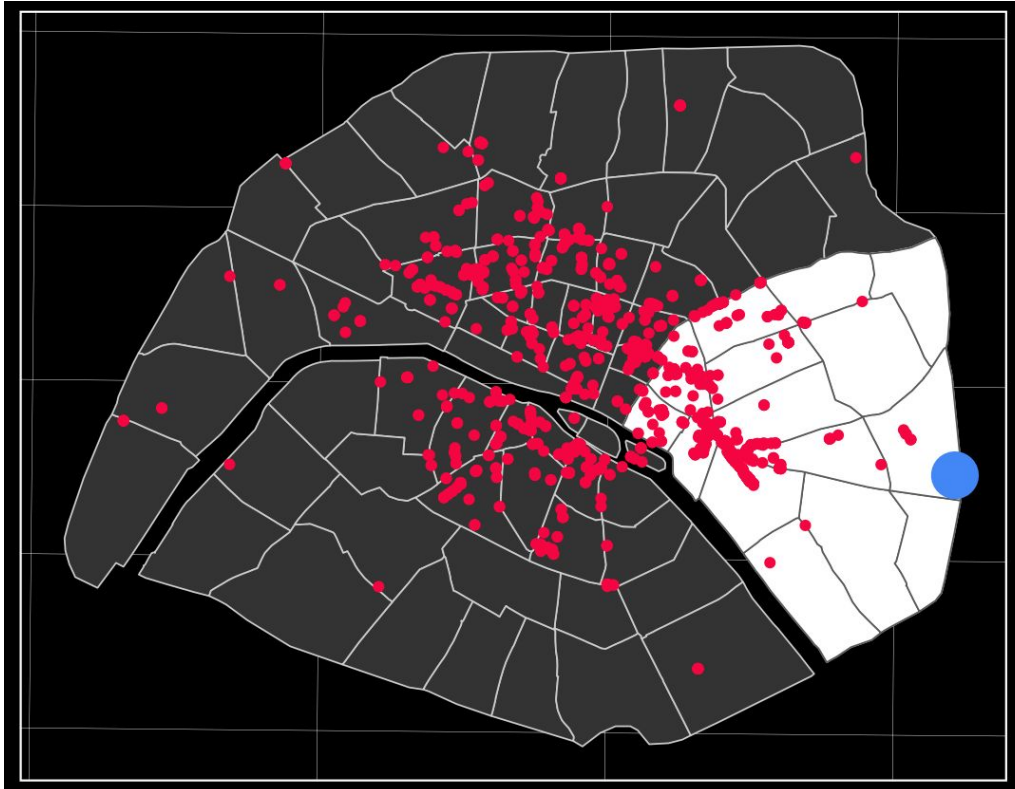
données V2, 2022-06-20, revues <https://doi.org/10.5281/zenodo.8388102>

Activités au sein des “Food stores”



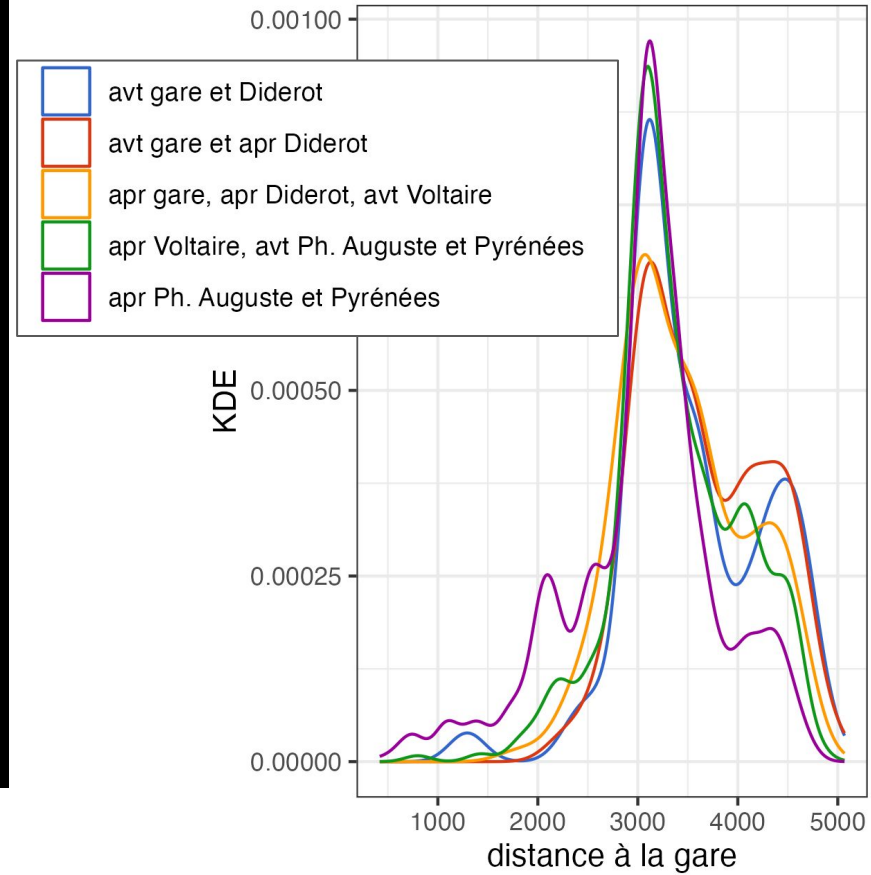
Des *Annuaires* dans le temps : snapshots d'espaces urbains

Ébénistes (1849-1853) et gare de Charonne-Marchandises



données V2, 2022-06-20

Ébénistes de l'Est parisien (1846-1885)



Des *Annuaire*s aux données longitudinales

Suivre des individus : cas des nourrisseurs

Données et graphe

entrées d'annuaires	7094
entrées avec relations dans le graphe	5625
cliques ayant au moins deux entrées	965
cliques sans homonymes	905

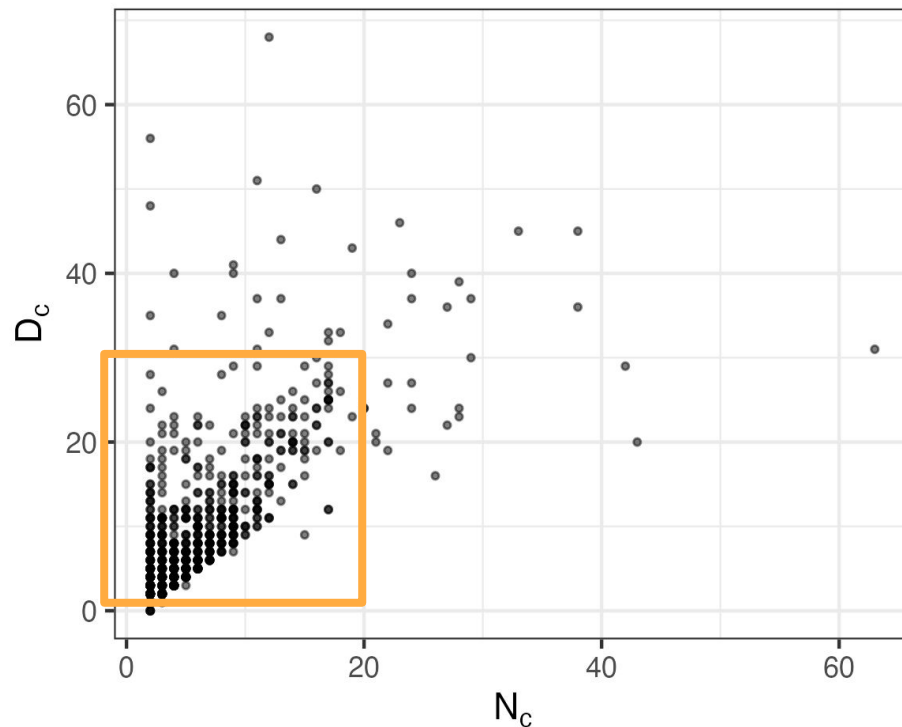
Cliques dans l'espace

nourrisseurs dont on peut étudier les changements spatiaux dans le temps	571
nourrisseurs au même emplacement	269
nourrisseurs ayant effectué au moins un mouvement dans l'espace	302

Mouvement

déménagements	91 %
expansion et/ou rétraction	9 %

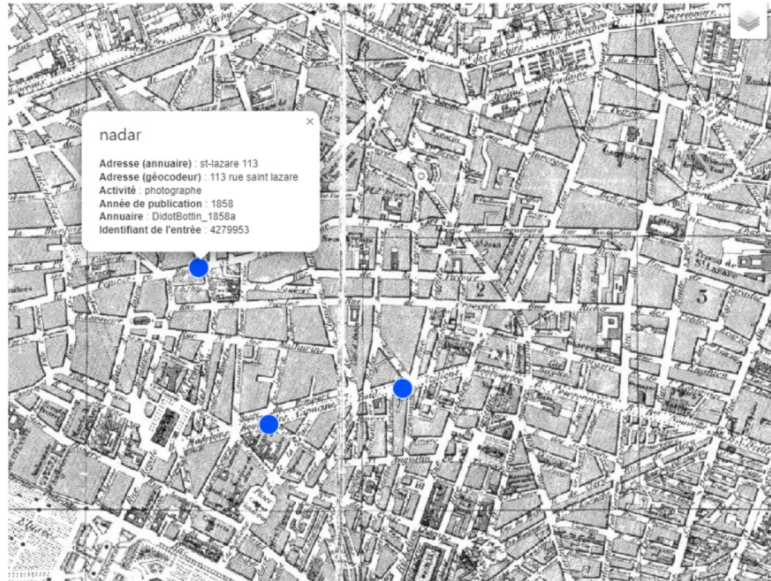
Détection d'homonymie : taille/durée des cliques



Des *Annuaire*s aux données longitudinales

Graphe géohistorique complexe : individus, successions & co-localisations des photographes

Déménagements du photographe Nadar



Filtres

Paramètres de recherche

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

Raison sociale

nadar

Activité

Mot-clé

Adresse

Adresse

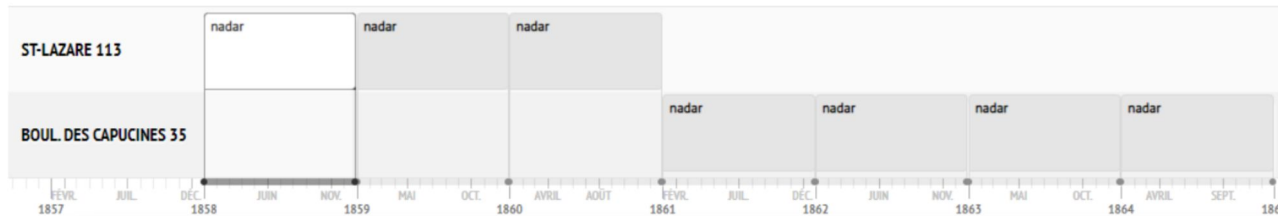
Valider

Filtrer par période

1830 1880

1790 1820 1850 1880 1910

S. Tual, N. Abadie, B. Duménieu, J. Chazalon, et E. Carlinet, « Création d'un graphe de connaissances géohistorique à partir d'annuaires du commerce parisien du 19^{ème} siècle: application aux métiers de la photographie », in 34^{es} Journées francophones d'Ingénierie des Connaissances (IC 2023) @ Plate-Forme Intelligence Artificielle (PFIA 2023), in IC2023 : 34^{es} Journées francophones d'Ingénierie des Connaissances. Strasbourg, France, juill. 2023. [En ligne]. Disponible sur: <https://hal.science/hal-04121643>



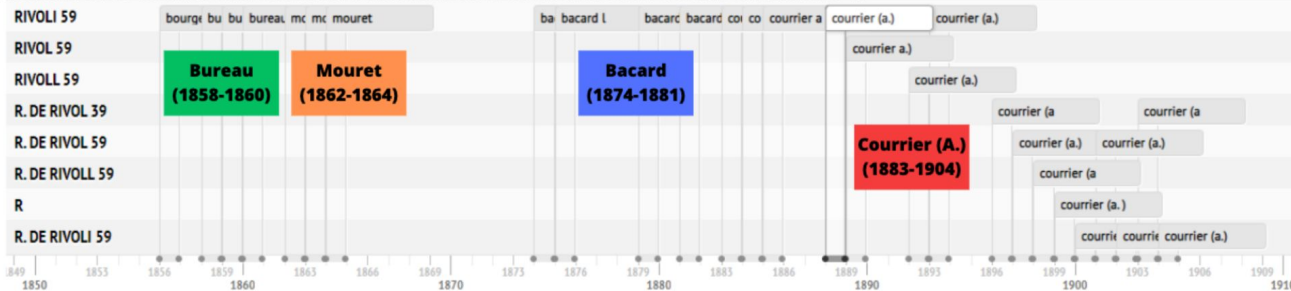
Des *Annuaire*s aux données longitudinales

Graphe géohistorique complexe : individus, successions & co-localisations des photographes

Phénomène de transmission probable d'un atelier entre photographes



S. Tual, N. Abadie, B. Duménieu, J. Chazalon, et E. Carlinet, « Création d'un graphe de connaissances géohistorique à partir d'annuaires du commerce parisien du 19^{ème} siècle: application aux métiers de la photographie », in 34es Journées francophones d'Ingénierie des Connaissances (IC 2023) @ Plate-Forme Intelligence Artificielle (PFIA 2023), in IC2023 : 34es Journées francophones d'Ingénierie des Connaissances. Strasbourg, France, juill. 2023. [En ligne]. Disponible sur: <https://hal.science/hal-04121643>



Conclusion



Profils d'utilisation, pistes d'analyse
et chantiers à poursuivre

Retour sur les annuaires et l'utilisation des données extraites

Retour sur la nature des Annuaires et de l'inscription d'une personne dans un tel ouvrage.

- Un produit éditorial, avec un intérêt pour les éditeurs de proposer de plus en plus d'adresses. Entreprise commerciale : inscriptions directes à partir de sources, souscriptions, et déclarations (payantes).
- Des gens mis dans une même liste, dans une même catégories par le simple fait de l'éditeur
- Des individus sont en représentation (autodéclaratif ou choix éditoriaux de catégorisation), souhaitent se distinguer ou apparaître dans une catégorie (ex. nouveautés)...

Données numériques des Annuaires :

- des "entrées" : c'est à dire des mentions d'"individus" ou de "collectifs" situés dans l'espace et associées à une activité.
- Question de l'alignement des informations : Rapprochement des entrées d'unités documentaires différentes pour suivre ces "individus" (mise en équivalence d'informations produites avec des logiques différentes). Problème classique en démographie et en histoire. On travaille sur des traces. des hypothèses mais aucune certitude
- Intérêts de raisonner avec les triplets. Appel à des sources complémentaire (base "réconciliation" wikipedia/data.bnf)

Paradoxe des l'analyse :

- Suivre ce qui ne bouge pas ou qui se ressemble fortement : repérer les "individus", des "boutiques", des catégories de professions
- Suivre ce qui bouge : trajectoires spatiales des individus (et étudier en quoi elles se distinguent ou non du mouvement général de la population et des activités professionnelles et de la structure spatiale)

Chantiers et dossiers à poursuivre :

Evaluation du corpus et des résultats de chaque étape de la chaîne

Finaliser l'étude critique du corpus des Annuaire et Almanach commerciaux Parisiens, notamment à l'aide des retours d'analyse des données.

Intégration et traitements des ouvrages qui participent du corpus à la suite de campagne de numérisations de sources. => V5

Nouvelle campagne d'annotation pour les pages annuaires, notamment pour les pages de fin de siècle avec les numéros de téléphone et arrondissements

Intégrer d'autres sources pour le géocodage et étudier la correspondance entre les différents numérotations (travail spécifique sur la Numérotation sectionnaire)

Cas d'études et confrontation avec d'autres sources pour comprendre ce que contiennent les annuaires (travaux présentés en nov. 2022 : Photographes, notaires, corps médical de Paris, gantiers, graveurs de cartes,...)

Pour conclure

- Une source numérique constituée "automatiquement"
- Transcription enrichie d'une source historique
- Une interprétation à minima et avec un lien à la source originale (page et boîte dans la page)
- Une source qui peut être citée : Id des entrées (par version)
- Une source pour des travaux de recherche pour plusieurs communautés scientifiques

Les données géohistoriques produites au cours du programme SoDUCo.
Évaluation d'une « source numérique »

MERCI